

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:24:09 ; Search time 54 Seconds
(Without alignments)
2267.728 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 919
Sequence: 1 MSKPYLFANRSEFPAALAA.....TGVKEEGNPILKHFITGTPF 919

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

A_Geneseq_101002:.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	919	100.0	919	21	AAV95987
2	788	85.7	889	21	AAV95988
3	11	1.2	576	21	AAV95821
4	11	1.2	578	21	AAV95820
5	9	1.0	262	21	AAV70579
6	9	1.0	851	14	AAAR41333
7	9	1.0	851	16	AAAR72077
8	9	1.0	851	17	AAW03166
9	9	1.0	851	21	AAAB12374
10	9	1.0	851	22	AAAB19963

11	9	1.0	851	23	AAE14649	Human STAT2 protei
12	9	1.0	851	23	AAE15173	Human Stat2 (signa
13	8	0.9	114	22	ABG25555	Novel human diago
14	8	0.9	331	23	ABBA7876	Listeria monocytog
15	8	0.9	371	22	ABG25558	Novel human diago
16	8	0.9	474	22	ABG17710	Novel human diago
17	8	0.9	582	19	AAW56318	Infectious Iarngog
18	8	0.9	1373	22	ABB62368	Drosophila melanog
19	7	0.8	29	14	AAAR4425	Mutant alpha-amyli
20	7	0.8	29	14	AAAR4426	Mutant alpha-amyli
21	7	0.8	29	14	AAAR4427	Mutant alpha-amyli
22	7	0.8	29	14	AAAR4428	Mutant alpha-amyli
23	7	0.8	34	16	AAAR4559	Protease Nexin-1 N
24	7	0.8	34	19	AAW51936	Human protease nex
25	7	0.8	39	23	AAU88768	Insulin/Insulin-11
26	7	0.8	41	7	AAU60714	Sequence of the fu
27	7	0.8	41	22	AAAB5508	Gene #5 associated
28	7	0.8	42	7	AAU60715	Sequence of the fu
29	7	0.8	51	21	AAV64841	Human 5' EST relat
30	7	0.8	51	22	AAU22430	Human cardiovascu
31	7	0.8	62	22	AAAB7002	Human Immune/Haema
32	7	0.8	64	20	AAV11419	Human 5' EST seque
33	7	0.8	64	22	AAAG77071	Human colon cancer
34	7	0.8	69	19	AAW75000	Human secreted pro
35	7	0.8	69	21	AAAG1031	Human secreted pro
36	7	0.8	69	21	AAAG1032	Human secreted pro
37	7	0.8	70	19	AAW74874	Human secreted pro
38	7	0.8	79	23	ABP05664	Human ORFX protein
39	7	0.8	82	22	ABAB16701	Human nervous syst
40	7	0.8	82	22	AAAG98900	E. coli growth and
41	7	0.8	88	21	AAAG1063	Zea mays protein f
42	7	0.8	95	17	AAW09286	Human mature Inter
43	7	0.8	95	18	AAW44113	Human Interferon t
44	7	0.8	102	21	AAAB54054	Human pancreatic c
45	7	0.8	104	21	AAAG57029	Arabidopsis thalia
46	7	0.8	106	21	AAAG37047	Arabidopsis thalia
47	7	0.8	106	21	AAAG47491	Arabidopsis thalia
48	7	0.8	106	21	AAAG47520	Arabidopsis thalia
49	7	0.8	106	22	AAW93109	Human digestive sy
50	7	0.8	106	22	AAW38625	Human colorectal c
51	7	0.8	107	21	AAAG57028	Arabidopsis thalia
52	7	0.8	108	21	AAAG57027	Arabidopsis thalia
53	7	0.8	111	21	AAAG40996	Arabidopsis thalia
54	7	0.8	112	22	AAAB65036	Zea mays protein f
55	7	0.8	124	21	AAAG1041	Gene #5 associated
56	7	0.8	125	22	AAAC01896	Zea mays protein f
57	7	0.8	129	23	AAAP01441	Human poly(amide
58	7	0.8	130	23	ABBA47564	Human ORFX protein
59	7	0.8	131	21	AAAG00735	Listeria monocytog
60	7	0.8	131	21	AAAG00734	Human secreted pro
61	7	0.8	136	21	AAAB40580	Human secreted pro
62	7	0.8	136	22	AAO51859	Human ORFX ORF344
63	7	0.8	136	22	ABG03607	Propionibacterium
64	7	0.8	136	22	AAO63373	Novel human diago
65	7	0.8	136	23	ABP33366	Human poly(amide
66	7	0.8	143	18	AAAB18163	Human ORF2339 prot
67	7	0.8	148	22	ABG22609	Mycobacterium tube
68	7	0.8	150	22	AAU87666	Novel human diago
69	7	0.8	150	22	AAU17564	Novel central nerv
70	7	0.8	151	22	ABG21654	Novel signal trans
71	7	0.8	154	21	AAAG1040	Novel human diago
72	7	0.8	156	21	AAAG37046	Zea mays protein f
73	7	0.8	156	21	AAAG37045	Arabidopsis thalia
74	7	0.8	156	21	AAAG47490	Arabidopsis thalia
75	7	0.8	156	22	AAU41076	Arabidopsis thalia
76	7	0.8	157	21	AAAG16420	Arabidopsis thalia
77	7	0.8	157	21	AAAG1061	Propionibacterium
78	7	0.8	159	19	AAW63034	Arabidopsis thalia
79	7	0.8	159	21	AAAB14325	Zea mays protein f
80	7	0.8	159	21	AAV91593	Mycobacterium tube
81	7	0.8	162	11	AAAR07326	Human secreted pro
82	7	0.8	162	22	AAU47747	Fusion of alpha am
83	7	0.8	162	22	AAU65590	Propionibacterium

84	7	0.8	163	18	AAW41812	N. tabacum strain
85	7	0.8	169	11	AAW07250	Protein specific t
86	7	0.8	170	21	AAW44967	zea mays protein f
87	7	0.8	180	22	AAU47350	Protonbacterium
88	7	0.8	187	21	AAW56274	Human secreted pro
89	7	0.8	189	22	AAU40254	Protonbacterium
90	7	0.8	192	21	AAW40892	zea mays protein f
91	7	0.8	194	21	AAW55688	Arabidopsis thalia
92	7	0.8	195	4	AAW30076	Sequence of bovine
93	7	0.8	195	11	AAW04541	Bovine interferon-
94	7	0.8	195	19	AAW73227	Bovine interferon-
95	7	0.8	195	20	AAW83898	Bovine interferon-
96	7	0.8	197	21	AAW55687	Arabidopsis thalia
97	7	0.8	199	21	AAW55686	Arabidopsis thalia
98	7	0.8	213	19	AAW80649	S. pneumoniae tran
99	7	0.8	214	22	AAU17391	Novel signal trans
100	7	0.8	215	22	AAU17391	P. syringae pv. de
101	7	0.8	217	23	AAW12600	Human DIRP polype
102	7	0.8	224	22	AAW12600	Novel human diagno
103	7	0.8	227	22	AAW12473	Novel human diagno
104	7	0.8	240	22	AAW62405	Human MBSP9 polype
105	7	0.8	255	23	AAW62405	Murine protein iso
106	7	0.8	261	10	AAW90438	Protein containing
107	7	0.8	261	21	AAW40566	Human ORFX ORF330
108	7	0.8	261	21	AAW94235	Human PRO1572 (UNQ
109	7	0.8	261	21	AAW94332	Human stomach prot
110	7	0.8	261	22	AAU29201	Human PPO polypept
111	7	0.8	261	22	AAW07051	Human gene 1 encod
112	7	0.8	261	22	AAW07073	Human gene 1 encod
113	7	0.8	261	22	AAW87584	Human PRO1572. Ho
114	7	0.8	261	22	AAW87584	Protein of the inv
115	7	0.8	261	22	AAW86181	Human albumin fusi
116	7	0.8	261	23	AAW65078	Human albumin fusi
117	7	0.8	261	23	AAW65079	Protein encoded by
118	7	0.8	261	23	AAW65091	Mouse Claudin-18 (
119	7	0.8	261	23	AAW65091	Lactococcus lactis
120	7	0.8	271	23	AAW55614	Human gene 1 encod
121	7	0.8	276	22	AAW07106	Human gene 1 encod
122	7	0.8	281	13	AAW28599	Human gene 1 encod
123	7	0.8	284	23	AAW895181	Human 158P1H4 spli
124	7	0.8	290	21	AAW09025	Arabidopsis thalia
125	7	0.8	295	22	AAW62034	Drosophila melanog
126	7	0.8	297	19	AAW29730	Endo-beta-1, 4-glu
127	7	0.8	297	19	AAW57777	Recombinant endo-1
128	7	0.8	303	22	AAW46002	Protonbacterium
129	7	0.8	308	22	AAW92563	C. glutamicum prote
130	7	0.8	310	23	AAW54218	Lactococcus lactis
131	7	0.8	311	22	AAW79756	Corynebacterium gl
132	7	0.8	311	23	AAW79756	Corynebacterium Cys
133	7	0.8	313	22	AAU24770	Human olfactory re
134	7	0.8	313	22	AAW71643	Human olfactory re
135	7	0.8	313	23	AAW95709	Human olfactory an
136	7	0.8	313	23	AAW85390	G-coupled olfactor
137	7	0.8	314	22	AAW03426	Novel human diagno
138	7	0.8	320	21	AAW09024	Arabidopsis thalia
139	7	0.8	320	22	AAW54656	Protonbacterium
140	7	0.8	321	22	AAW71629	Drosophila melanog
141	7	0.8	321	22	AAW71629	E. coli cellular p
142	7	0.8	323	22	AAU34728	Salmonella typhi c
143	7	0.8	323	23	AAW01873	Isoprenoid related
144	7	0.8	323	23	AAW01873	Arabidopsis thalia
145	7	0.8	325	21	AAW09023	Arabidopsis thalia
146	7	0.8	327	21	AAW22231	Plant SDF encoded
147	7	0.8	328	21	AAW24871	Novel human diagno
148	7	0.8	329	21	AAW24871	Novel human diagno
149	7	0.8	334	21	AAW6181	Arabidopsis thalia
150	7	0.8	338	21	AAW22230	Drosophila thalia
151	7	0.8	339	22	AAW67975	Pseudomonas aerugi
152	7	0.8	343	22	AAU36294	Human PRO5773 poly
153	7	0.8	343	22	AAU12252	S. epidermidis ope
154	7	0.8	343	22	AAW82365	S. epidermidis ope
155	7	0.8	344	21	AAW24870	Plant SDF encoded
156	7	0.8	347	21	AAW22229	Arabidopsis thalia
157	7	0.8	349	18	AAW08101	G-protein coupled
158	7	0.8	349	20	AAV17749	Human HIBPF51 rece
159	7	0.8	349	21	AAW90621	Human G protein-co
160	7	0.8	349	21	AAW90655	Human mutant G pro
161	7	0.8	349	21	AAW70341	Human G protein-co
162	7	0.8	350	20	AAW02170	A. Staphylococcus a
163	7	0.8	350	22	AAU34216	Staphylococcus aur
164	7	0.8	350	22	AAU36879	Staphylococcus aur
165	7	0.8	350	22	AAW30030	Novel human diagno
166	7	0.8	351	22	AAW30030	Novel human diagno
167	7	0.8	351	22	AAW30030	Novel human diagno
168	7	0.8	352	22	AAW30030	Novel human diagno
169	7	0.8	352	22	AAW30030	Novel human diagno
170	7	0.8	352	22	AAW30030	Novel human diagno
171	7	0.8	352	22	AAW30030	Novel human diagno
172	7	0.8	352	22	AAW30030	Novel human diagno
173	7	0.8	352	22	AAW30030	Novel human diagno
174	7	0.8	352	22	AAW30030	Novel human diagno
175	7	0.8	352	22	AAW30030	Novel human diagno
176	7	0.8	352	22	AAW30030	Novel human diagno
177	7	0.8	352	22	AAW30030	Novel human diagno
178	7	0.8	352	22	AAW30030	Novel human diagno
179	7	0.8	352	22	AAW30030	Novel human diagno
180	7	0.8	352	22	AAW30030	Novel human diagno
181	7	0.8	352	22	AAW30030	Novel human diagno
182	7	0.8	352	22	AAW30030	Novel human diagno
183	7	0.8	352	22	AAW30030	Novel human diagno
184	7	0.8	352	22	AAW30030	Novel human diagno
185	7	0.8	352	22	AAW30030	Novel human diagno
186	7	0.8	352	22	AAW30030	Novel human diagno
187	7	0.8	352	22	AAW30030	Novel human diagno
188	7	0.8	352	22	AAW30030	Novel human diagno
189	7	0.8	352	22	AAW30030	Novel human diagno
190	7	0.8	352	22	AAW30030	Novel human diagno
191	7	0.8	352	22	AAW30030	Novel human diagno
192	7	0.8	352	22	AAW30030	Novel human diagno
193	7	0.8	352	22	AAW30030	Novel human diagno
194	7	0.8	352	22	AAW30030	Novel human diagno
195	7	0.8	352	22	AAW30030	Novel human diagno
196	7	0.8	352	22	AAW30030	Novel human diagno
197	7	0.8	352	22	AAW30030	Novel human diagno
198	7	0.8	352	22	AAW30030	Novel human diagno
199	7	0.8	352	22	AAW30030	Novel human diagno
200	7	0.8	352	22	AAW30030	Novel human diagno
201	7	0.8	352	22	AAW30030	Novel human diagno
202	7	0.8	352	22	AAW30030	Novel human diagno
203	7	0.8	352	22	AAW30030	Novel human diagno
204	7	0.8	352	22	AAW30030	Novel human diagno
205	7	0.8	352	22	AAW30030	Novel human diagno
206	7	0.8	352	22	AAW30030	Novel human diagno
207	7	0.8	352	22	AAW30030	Novel human diagno
208	7	0.8	352	22	AAW30030	Novel human diagno
209	7	0.8	352	22	AAW30030	Novel human diagno
210	7	0.8	352	22	AAW30030	Novel human diagno
211	7	0.8	352	22	AAW30030	Novel human diagno
212	7	0.8	352	22	AAW30030	Novel human diagno
213	7	0.8	352	22	AAW30030	Novel human diagno
214	7	0.8	352	22	AAW30030	Novel human diagno
215	7	0.8	352	22	AAW30030	Novel human diagno
216	7	0.8	352	22	AAW30030	Novel human diagno
217	7	0.8	352	22	AAW30030	Novel human diagno
218	7	0.8	352	22	AAW30030	Novel human diagno
219	7	0.8	352	22	AAW30030	Novel human diagno
220	7	0.8	352	22	AAW30030	Novel human diagno
221	7	0.8	352	22	AAW30030	Novel human diagno
222	7	0.8	352	22	AAW30030	Novel human diagno
223	7	0.8	352	22	AAW30030	Novel human diagno
224	7	0.8	352	22	AAW30030	Novel human diagno
225	7	0.8	352	22	AAW30030	Novel human diagno
226	7	0.8	352	22	AAW30030	Novel human diagno
227	7	0.8	352	22	AAW30030	Novel human diagno
228	7	0.8	352	22	AAW30030	Novel human diagno
229	7	0.8	352	22	AAW30030	Novel human diagno

230	7	0.8	398	23	ABR09439	Serpin domain prot
231	7	0.8	400	22	ABG02720	Neuron-associated
232	7	0.8	408	22	AAB30705	Membrane-bound pro
233	7	0.8	412	20	AAV35532	Human p90 polypept
234	7	0.8	419	22	AAB64997	Human polypeptide
235	7	0.8	420	23	ABR76313	Human PRO1136. HO
236	7	0.8	433	22	ABR67364	Human PRO1136 (UNQ
237	7	0.8	433	23	ABR61037	Plasmodium regula
238	7	0.8	434	17	AAR96419	Novel human diago
239	7	0.8	434	21	AAB12820	Drosophila melanog
240	7	0.8	441	21	AAG36179	Pinus radiata cell
241	7	0.8	443	21	AAV91592	S. fradiae tylosin
242	7	0.8	443	22	ABG26155	Human protein cont
243	7	0.8	443	23	ABR93726	Drosophila melanog
244	7	0.8	449	21	AAB24869	Novel signal trans
245	7	0.8	458	23	AAB81986	Candida tropicalis
246	7	0.8	468	22	ABG00728	Drosophila melanog
247	7	0.8	469	22	AAM40870	Human secreted pro
248	7	0.8	474	22	AAV89794	Aquifex pyrophilus
249	7	0.8	480	22	AAU33686	Novel human diago
250	7	0.8	481	22	ABR70673	Human tumour-assoc
251	7	0.8	488	19	AAV49025	Alpha-amy-lase-cell
252	7	0.8	491	22	AAU87389	Novel central nerv
253	7	0.8	491	22	AAU17172	Novel signal trans
254	7	0.8	491	22	AAU07844	Novel human serine
255	7	0.8	494	22	AAV92620	C. glutamicum prote
256	7	0.8	494	22	AAV78986	Glucocanase gene
257	7	0.8	497	8	AAV70572	Alpha-amy-lase gene
258	7	0.8	511	8	AAV70339	B. licheniformis a
259	7	0.8	511	19	AAV39741	An alpha-amy-lase p
260	7	0.8	511	20	AAV01582	Novel human diago
261	7	0.8	511	20	AAV01583	Novel human diago
262	7	0.8	512	8	AAV70753	B. licheniformis a
263	7	0.8	512	18	AAV10325	Wild type alpha-am
264	7	0.8	512	20	AAV22181	B. licheniformis a
265	7	0.8	512	20	AAV00769	Alpha-amy-lase prot
266	7	0.8	512	20	AAV73509	B. licheniformis a
267	7	0.8	512	22	AAV65875	B. licheniformis a
268	7	0.8	512	22	AAV65877	B. licheniformis a
269	7	0.8	512	22	AAU07097	Bacillus lichenifo
270	7	0.8	512	22	AAU07098	Bacillus lichenifo
271	7	0.8	512	22	AAU07099	Bacillus lichenifo
272	7	0.8	512	22	AAU07100	Bacillus lichenifo
273	7	0.8	512	22	AAU07549	Bacillus lichenifo
274	7	0.8	512	22	AAU07550	Bacillus lichenifo
275	7	0.8	512	22	AAU07551	Bacillus lichenifo
276	7	0.8	512	22	AAU07552	Bacillus lichenifo
277	7	0.8	512	22	AAU07553	Bacillus lichenifo
278	7	0.8	512	22	AAU07554	Bacillus lichenifo
279	7	0.8	512	22	AAU07555	Bacillus lichenifo
280	7	0.8	512	22	AAU07556	Bacillus lichenifo
281	7	0.8	513	22	AAU31141	Novel human secret
282	7	0.8	516	20	AAV93433	A. thaliana ELV pr
283	7	0.8	517	18	AAV34204	Streptomyces efllu
284	7	0.8	517	19	AAV53605	Streptomycetes rose
285	7	0.8	520	11	AAV04571	ORF3 product from
286	7	0.8	524	22	AAV62989	Drosophila melanog
287	7	0.8	524	22	ABR62990	Drosophila melanog
288	7	0.8	534	20	AAV29133	Amino acid sequenc
289	7	0.8	548	22	ABR62279	Drosophila melanog
290	7	0.8	551	21	AAV16422	Pinus radiata 4CL
291	7	0.8	560	18	AAV23602	Alpha-amy-lase-cell
292	7	0.8	560	18	AAV22522	Alpha-amy-lase-cell
293	7	0.8	560	22	AAV88599	Human hydrophobic
294	7	0.8	567	22	ABR68354	Drosophila melanog
295	7	0.8	567	22	AAV94236	Human protein sequ
296	7	0.8	574	19	AAV57434	Terminyl-linker-CB
297	7	0.8	593	23	ABR30097	Streptococcus poly
298	7	0.8	597	22	ABG18227	Novel human diago
299	7	0.8	599	22	AAV91428	C. glutamicum prote
300	7	0.8	599	22	AAV76686	Corynebacterium gl
301	7	0.8	624	17	AAV77674	Glucocanase from
302	7	0.8	630	18	AAV23603	Alpha-amy-lase-Cena

Alpha-amy-lase-Cena
Neuron-associated
Membrane-bound pro
Human p90 polypept
Human polypeptide
Human PRO1136. HO
Human PRO1136 (UNQ
Plasmodium regula
Novel human diago
Drosophila melanog
Pinus radiata cell
S. fradiae tylosin
Human protein cont
Drosophila melanog
Novel signal trans
Candida tropicalis
Drosophila melanog
Human secreted pro
Aquifex pyrophilus
Novel human diago
Human tumour-assoc
Alpha-amy-lase-cell
Terminyl-linker-CB
Sequence of Crypto
Drosophila melanog
Herbically activ
Lung small cell ca
Novel human diago
Novel human diago
A Bacillus pectate
Human tumour-assoc
HIV-1 non-subtype
Eucalyptus grandis
Arabidopsis thalia
Arabidopsis thalia
Eucalyptus grandis
Drosophila melanog
Drosophila melanog
Rat GRP2. Ratius
Arabidopsis thalia
p110. AAR43341;
Human p110. Homo
Ptdins 3-kinase 11
Human p110alpha is
Human protein sequ
Human Cl channel n
Drosophila melanog
S-layer protein en
Bacillus stearothhe
B. stearothemophi
B. stearothemophi
Novel human diago
Chicken Nr-CAM pro
Drosophila melanog
Drosophila melanog
Novel human diago
Drosophila melanog
P. horikoshii PHBN
Drosophila melanog
Novel human secret
C. glutamicum prote
Subsequence which
Hepatitis B virus-
peptide fragment (p
HBV pol peptide (p
Hepatitis B virus
Synthetic polypept
Immunogenic peptid
Immunogenic peptid
MHC class I assoc
Claudin cell adhes

376	6	0.7	9	22	AA93792	Hepatitis B virus
377	6	0.7	9	22	AA95942	MHC class-I associ
378	6	0.7	9	22	AA95942	CAR sequence, a ca
379	6	0.7	10	15	AA93863	Antigen fragment 1
380	6	0.7	10	21	AA948614	CaMk I Ser/Thr kin
381	6	0.7	10	22	AA96396	Human prostate ant
382	6	0.7	10	22	AA96420	Human leukocyte an
383	6	0.7	10	22	AA96420	Human leukocyte an
384	6	0.7	10	22	AA96420	Human leukocyte an
385	6	0.7	10	22	AA96420	Human complementar
386	6	0.7	11	21	AA97991	Prostatic acid pho
387	6	0.7	11	21	AA97991	HIV-2 detection va
388	6	0.7	11	22	AA96419	Human leukocyte an
389	6	0.7	11	22	AA96419	Human leukocyte an
390	6	0.7	12	21	AA939602	Anti-IL12 antibody
391	6	0.7	12	21	AA939602	Anti-IL12 antibody
392	6	0.7	12	21	AA939602	Anti-IL12 antibody
393	6	0.7	13	12	AA961754	Human KRP1 tryptic
394	6	0.7	13	12	AA961754	Peptide with affini
395	6	0.7	13	12	AA961754	Peptide with affini
396	6	0.7	13	12	AA961754	Peptide with affini
397	6	0.7	13	12	AA961754	Peptide with affini
398	6	0.7	13	12	AA961754	Peptide with affini
399	6	0.7	13	12	AA961754	Peptide with affini
400	6	0.7	13	12	AA961754	Peptide with affini
401	6	0.7	14	14	AA940123	Schizophrenia-asso
402	6	0.7	14	18	AA936074	Delta subunit prot
403	6	0.7	15	15	AA94856	E. coli DNA polyme
404	6	0.7	15	20	AA929718	Mycobacterium tube
405	6	0.7	15	20	AA929718	Lecithin:cholester
406	6	0.7	15	20	AA929718	Lecithin:cholester
407	6	0.7	15	20	AA929718	Lecithin:cholester
408	6	0.7	15	20	AA929718	Lecithin:cholester
409	6	0.7	15	23	AA918710	Major histocompati
410	6	0.7	16	21	AA918710	Human zinc finger
411	6	0.7	16	21	AA918710	Pro-TNF peptide fr
412	6	0.7	16	21	AA918710	HIV-2 detection va
413	6	0.7	16	21	AA918710	Moraxella catarin
414	6	0.7	16	23	AA932417	Topacco mosaic vir
415	6	0.7	17	18	AA918709	Major histocompati
416	6	0.7	17	18	AA918709	Peptide derived fr
417	6	0.7	17	18	AA918709	HIV infection dete
418	6	0.7	17	22	AA918709	Novel human diapo
419	6	0.7	17	22	AA918709	Major histocompati
420	6	0.7	18	21	AA918709	HIV detection rela
421	6	0.7	18	21	AA918709	HIV detection rela
422	6	0.7	18	21	AA918709	HIV detection rela
423	6	0.7	18	21	AA918709	HIV detection rela
424	6	0.7	19	20	AA929096	Human mammary asso
425	6	0.7	22	21	AA91061	L. monocytogenes s
426	6	0.7	22	21	AA91061	HIV infection dete
427	6	0.7	22	21	AA91061	HIV infection dete
428	6	0.7	22	21	AA91061	HIV infection dete
429	6	0.7	22	21	AA91061	HIV infection dete
430	6	0.7	22	21	AA91061	HIV infection dete
431	6	0.7	22	21	AA91061	HIV infection dete
432	6	0.7	22	21	AA91061	HIV infection dete
433	6	0.7	22	21	AA91061	HIV infection dete
434	6	0.7	22	21	AA91061	HIV infection dete
435	6	0.7	22	21	AA91061	HIV infection dete
436	6	0.7	22	21	AA91061	HIV infection dete
437	6	0.7	22	21	AA91061	HIV infection dete
438	6	0.7	22	21	AA91061	HIV infection dete
439	6	0.7	22	21	AA91061	HIV infection dete
440	6	0.7	22	21	AA91061	HIV infection dete
441	6	0.7	22	21	AA91061	HIV infection dete
442	6	0.7	22	21	AA91061	HIV infection dete
443	6	0.7	22	21	AA91061	HIV infection dete
444	6	0.7	22	21	AA91061	HIV infection dete
445	6	0.7	22	21	AA91061	HIV infection dete
446	6	0.7	22	21	AA91061	HIV infection dete
447	6	0.7	22	21	AA91061	HIV infection dete
448	6	0.7	22	21	AA91061	HIV infection dete

522	6	0.7	39	23	AAU88751	Insulin/insulin-11	595	6	0.7	50	22	AAU33786	Peptide #7823 enco
523	6	0.7	39	23	AAU88761	Insulin/insulin-11	596	6	0.7	50	23	ABG43459	Human peptide enco
524	6	0.7	39	23	AAU88765	Insulin/insulin-11	597	6	0.7	50	23	AAU90802	Insulin/insulin-11
525	6	0.7	39	23	AAU88771	Insulin/insulin-11	598	6	0.7	52	21	AAU09489	Arabidopsis thalia
526	6	0.7	39	23	AAU88773	Insulin/insulin-11	599	6	0.7	52	21	AAU02982	Human secreted pro
527	6	0.7	39	23	AAU90800	Insulin/insulin-11	600	6	0.7	52	22	AAU51595	Propionibacterium
528	6	0.7	40	19	AAU78737	Leader peptide seq	601	6	0.7	53	21	AAU50046	Arabidopsis thalia
529	6	0.7	40	19	AAU78741	Leader peptide seq	602	6	0.7	54	21	AAU50045	Arabidopsis thalia
530	6	0.7	40	20	AAU72373	Xylanase II secret	603	6	0.7	54	21	AAU53245	Arabidopsis thalia
531	6	0.7	40	20	AAU12038	Human 5' EST secret	604	6	0.7	54	22	AAU39729	Propionibacterium
532	6	0.7	40	21	AAU58409	Lung cancer associ	605	6	0.7	54	22	AAU42051	Propionibacterium
533	6	0.7	40	21	AAU79833	HIV infection dete	606	6	0.7	54	22	AAU63984	Propionibacterium
534	6	0.7	40	22	ABU36101	Peptide #3607 enco	607	6	0.7	54	22	AAU66104	Propionibacterium
535	6	0.7	40	22	ABU21491	Protein #3490 enco	608	6	0.7	54	23	ABU51854	Human colon specif
536	6	0.7	40	22	AAU56882	Human brain expres	609	6	0.7	54	23	ABU05061	Human ORFX protein
537	6	0.7	40	22	AAU69272	Human bone marrow	610	6	0.7	55	22	AAU43640	Propionibacterium
538	6	0.7	40	22	AAU17102	Peptide #3536 enco	611	6	0.7	55	22	AAU47749	Propionibacterium
539	6	0.7	40	22	AAU29592	Peptide #3629 enco	612	6	0.7	55	22	AAU49783	Propionibacterium
540	6	0.7	40	22	AAU74059	Human colon cancer	613	6	0.7	55	22	AAU22774	Human prostate can
541	6	0.7	40	22	AAU74403	Vector pXB61-TV en	614	6	0.7	55	22	AAU94755	Human reproductive
542	6	0.7	40	23	ABU38885	Human peptide enco	615	6	0.7	56	22	AAU04653	Haem protein relat
543	6	0.7	40	23	AAU88752	Insulin/insulin-11	616	6	0.7	57	22	AAU61850	Propionibacterium
544	6	0.7	41	16	AAU80538	Moloney murine leu	617	6	0.7	57	22	AAU63889	Propionibacterium
545	6	0.7	41	16	AAU76597	MOuV gag matrix p	618	6	0.7	57	22	AAU92691	Human digestive sy
546	6	0.7	41	17	AAU88178	Leader sequence fo	619	6	0.7	57	22	AAU23824	Human EST encoded
547	6	0.7	41	19	AAU78727	Leader peptide seq	620	6	0.7	58	22	AAU59603	Propionibacterium
548	6	0.7	41	19	AAU78735	Leader peptide seq	621	6	0.7	58	22	AAU59909	Propionibacterium
549	6	0.7	41	19	AAU78738	Leader peptide seq	622	6	0.7	58	22	AAU85462	Human immune/haema
550	6	0.7	41	19	AAU78739	Leader peptide seq	623	6	0.7	59	19	AAU76664	Elapidae modified
551	6	0.7	41	21	AAU56118	Human secreted pro	624	6	0.7	59	19	AAU76666	Elapidae modified
552	6	0.7	41	21	AAU27179	PIV3 L polymerase	625	6	0.7	59	21	AAU57574	Arabidopsis thalia
553	6	0.7	41	21	AAU27197	PIV3 L polymerase	626	6	0.7	59	21	AAU59560	Arabidopsis thalia
554	6	0.7	41	21	AAU21570	Cone snail alpha-c	627	6	0.7	59	22	AAU61825	Propionibacterium
555	6	0.7	42	19	AAU78719	Leader peptide seq	628	6	0.7	59	23	ABU33467	Human ORF2440 prot
556	6	0.7	42	21	AAU78722	Leader peptide seq	629	6	0.7	60	21	AAU15112	Arabidopsis thalia
557	6	0.7	42	21	AAU79774	HIV detection rela	630	6	0.7	60	22	AAU39247	Propionibacterium
558	6	0.7	42	21	AAU79840	HIV infection dete	631	6	0.7	60	22	AAU64705	Propionibacterium
559	6	0.7	42	21	ABU02387	Novel human diagno	632	6	0.7	61	21	AAU07969	Arabidopsis thalia
560	6	0.7	43	18	AAU19135	LA19 leader peptide	633	6	0.7	61	21	AAU12116	Arabidopsis thalia
561	6	0.7	43	19	AAU78726	Leader peptide seq	634	6	0.7	61	21	AAU00892	Human secreted pro
562	6	0.7	43	19	AAU78717	Leader peptide seq	635	6	0.7	61	22	ABU59394	Drosophila melanog
563	6	0.7	43	19	AAU78718	Leader peptide seq	636	6	0.7	61	22	AAU88413	Human immune/haema
564	6	0.7	43	19	AAU69152	Yeast expression s	637	6	0.7	62	22	ABU27472	Streptococcus poly
565	6	0.7	43	19	AAU69151	Yeast expression s	638	6	0.7	62	22	ABU43391	Peptide #10897 enc
566	6	0.7	43	20	AAU13987	P60 repeat sequenc	639	6	0.7	62	22	ABU16601	Human nervous syst
567	6	0.7	43	23	AAU90779	Insulin/insulin-11	640	6	0.7	62	22	ABU17995	Human nervous syst
568	6	0.7	44	22	AAU89297	Human immune/haema	641	6	0.7	62	22	ABU26365	Protein #8364 enco
569	6	0.7	44	22	AAU10101	Human polypeptide	642	6	0.7	62	22	AAU64315	Human brain expres
570	6	0.7	46	21	AAU79813	HIV infection dete	643	6	0.7	62	22	AAU77143	Human bone marrow
571	6	0.7	46	21	AAU79815	HIV infection dete	644	6	0.7	62	22	AAU83881	Human immune/haema
572	6	0.7	46	22	ABU31712	Peptide #4363 enco	645	6	0.7	62	22	AAU21077	Peptide #7511 enco
573	6	0.7	46	22	ABU36932	Peptide #4438 enco	646	6	0.7	62	22	AAU37272	Peptide #11309 enc
574	6	0.7	46	22	ABU40181	Protein #7687 enco	647	6	0.7	62	23	ABU46154	Human peptide enco
575	6	0.7	46	22	ABU24627	Human bone marrow	648	6	0.7	62	23	ABU03006	Human ORFX protein
576	6	0.7	46	22	AAU60949	Human brain expres	649	6	0.7	63	22	AAU39445	Propionibacterium
577	6	0.7	46	22	AAU73642	Human bone marrow	650	6	0.7	63	22	AAU34679	Propionibacterium
578	6	0.7	46	22	AAU20047	Peptide #6481 enco	651	6	0.7	63	22	AAU51588	Propionibacterium
579	6	0.7	46	22	AAU30414	Peptide #4451 enco	652	6	0.7	63	23	ABU32712	Human cadherin-11k
580	6	0.7	46	22	AAU33833	Peptide #7870 enco	653	6	0.7	63	23	ABU11177	Human ORFX protein
581	6	0.7	46	23	ABU39710	Human peptide enco	654	6	0.7	64	21	AAU65012	Human 5' EST relat
582	6	0.7	46	23	ABU43514	Human peptide enco	655	6	0.7	64	22	AAU44953	Propionibacterium
583	6	0.7	47	22	AAU86352	Human immune/haema	656	6	0.7	64	22	AAU92782	Human digestive sy
584	6	0.7	47	22	AAU87332	Human immune/haema	657	6	0.7	65	22	AAU64709	Propionibacterium
585	6	0.7	48	20	AAU73915	Human prostate tum	658	6	0.7	65	22	AAU09187	Human polypeptide
586	6	0.7	48	22	AAU59509	Human brain expres	659	6	0.7	65	23	ABU04894	Human ORFX protein
587	6	0.7	49	16	AAU77796	Transactivation do	660	6	0.7	66	21	AAU54237	Human pancreatic c
588	6	0.7	49	16	AAU77797	Transactivation do	661	6	0.7	66	22	AAU43022	Propionibacterium
589	6	0.7	49	21	AAU34649	Arabidopsis thalia	662	6	0.7	66	22	AAU45044	Propionibacterium
590	6	0.7	50	22	AAU55386	Propionibacterium	663	6	0.7	66	22	AAU62876	Propionibacterium
591	6	0.7	50	22	AAU40144	Peptide #7650 enco	664	6	0.7	66	22	AAU31701	Novel human secret
592	6	0.7	50	22	AAU60907	Human brain expres	665	6	0.7	66	22	AAU07819	Human polypeptide
593	6	0.7	50	22	AAU73591	Human bone marrow	666	6	0.7	66	23	ABU00389	Human ORFX protein
594	6	0.7	50	22	AAU13825	Human polypeptide	667	6	0.7	66	23	ABU05956	Human ORFX protein

658	6	0.7	67	18	AAW27986	Staphylococcus aut
659	6	0.7	67	22	ABG23990	Novel human diagno
670	6	0.7	67	22	ABG29430	Novel human diagno
671	6	0.7	67	22	AAU020806	Human novel foetal
672	6	0.7	68	12	AAAR10107	Kistlin. Syntheti
673	6	0.7	68	13	AAAR25191	PIA from Agkistrod
674	6	0.7	68	15	AAAR33943	Disintegrin peptid
675	6	0.7	68	21	AAAY55657	M. genitalium ycfB
676	6	0.7	68	22	AAAB95638	Human testicular a
677	6	0.7	68	22	AAAG41177	Human reproductive
678	6	0.7	68	22	AAAM88847	Human immune/haema
679	6	0.7	68	23	ABP38811	Staphylococcus epi
680	6	0.7	68	23	ABP06283	Human OREX protein
681	6	0.7	69	21	AAAG58275	Arabidopsis thalia
682	6	0.7	69	22	ABAB69649	Drosophila melanog
683	6	0.7	69	22	AAUA45102	Proplionibacterium
684	6	0.7	69	22	ABG29142	Novel human diagno
685	6	0.7	69	22	ABAB28290	Human peptide #941
686	6	0.7	69	22	ABAB33470	Peptide #976 encod
687	6	0.7	69	22	ABAB38759	Peptide #6265 enco
688	6	0.7	69	22	ABAB18930	Protein #929 encod
689	6	0.7	69	22	ABAB23803	Human peptide #5802
690	6	0.7	69	22	AAAB54247	Human brain expres
691	6	0.7	69	22	AAAB59398	Human bone marrow
692	6	0.7	69	22	AAAB66640	Human bone marrow
693	6	0.7	69	22	AAAB71953	Human bone marrow
694	6	0.7	69	22	AAAB14513	Peptide #947 encod
695	6	0.7	69	22	AAAB26930	Peptide #967 encod
696	6	0.7	69	22	AAAB32232	Peptide #6269 enco
697	6	0.7	69	22	AAAB02239	Peptide #921 encod
698	6	0.7	69	23	ABG36298	Human peptide enco
699	6	0.7	69	23	ABG41768	Human peptide enco
700	6	0.7	69	23	ABP00313	Human OREX protein
701	6	0.7	70	22	ABAB70871	Drosophila melanog
702	6	0.7	70	22	AAAB60025	Proplionibacterium
703	6	0.7	70	23	ABP01387	Human OREX protein
704	6	0.7	70	23	ABP06857	Human OREX protein
705	6	0.7	71	14	AAAR31609	Encoded by chicken
706	6	0.7	71	14	AAAR32874	gry-1 protein. Ne
707	6	0.7	71	21	AAAG12319	Zea mays protein f
708	6	0.7	71	21	AAAG12319	Proplionibacterium
709	6	0.7	71	22	AAU43100	Peptide #4073 enco
710	6	0.7	71	22	AAU59712	Peptide #3801 enco
711	6	0.7	71	22	ABAB31372	Peptide #4084 enco
712	6	0.7	71	22	ABAB36295	Protein #3919 enco
713	6	0.7	71	22	ABAB36578	Human brain expres
714	6	0.7	71	22	ABAB21920	Human brain expres
715	6	0.7	71	22	AAAM57066	Human brain expres
716	6	0.7	71	22	AAAM57343	Human bone marrow
717	6	0.7	71	22	AAAM69741	Human bone marrow
718	6	0.7	71	22	AAAM77286	Human bone marrow
719	6	0.7	71	22	AAAM17558	Peptide #3992 enco
720	6	0.7	71	22	AAAM30082	Peptide #4119 enco
721	6	0.7	71	22	AAAM37438	Peptide #11475 enc
722	6	0.7	71	23	AAAM05225	Peptide #3907 enco
723	6	0.7	71	23	ABJ03767	Human ovary specif
724	6	0.7	71	23	ABG39077	Human peptide enco
725	6	0.7	71	23	ABG39365	Human peptide enco
726	6	0.7	71	23	ABG46301	Human peptide enco
727	6	0.7	72	21	ABP31491	Human OREX464 prote
728	6	0.7	72	21	AAAG54855	Arabidopsis thalia
729	6	0.7	72	21	AAAG03246	Human secreted pro
730	6	0.7	72	22	AAU655522	Proplionibacterium
731	6	0.7	72	22	AAAM85279	Human immune/haema
732	6	0.7	72	22	AAU02011	Human EST encoded
733	6	0.7	72	23	ABP34690	Human secreted pro
734	6	0.7	72	23	ABP01746	Human synthase-11k
735	6	0.7	73	22	AAU050111	Human OREX protein
736	6	0.7	73	22	ABAB27958	Proplionibacterium
737	6	0.7	73	22	ABAB33130	Human peptide #609
738	6	0.7	73	22	ABAB18595	Peptide #636 encod
739	6	0.7	73	22	AAAM53926	Protein #594 encod
740	6	0.7	73	22	AAAM66314	Human brain expres
						Human bone marrow

814	6	0.7	87	21	AA612343	Zea mays protein f
815	6	0.7	87	21	AA633517	Arabidopsis thalia
816	6	0.7	87	22	AA657419	Propionibacterium
817	6	0.7	87	22	AB611467	Novel human diagno
818	6	0.7	87	22	AB630017	Novel human diagno
819	6	0.7	88	22	AA657235	Propionibacterium
820	6	0.7	88	22	AB612946	Novel human diagno
821	6	0.7	88	22	AA685817	Human immune/haema
822	6	0.7	88	23	AB635030	Human structural p
823	6	0.7	88	23	AB648668	Listeria monocytog
824	6	0.7	89	21	AA638255	Human secreted pro
825	6	0.7	89	21	AB638257	Human secreted pro
826	6	0.7	89	22	AB611062	Novel human diagno
827	6	0.7	89	22	AB606314	Human OREX protein
828	6	0.7	90	19	AA680333	Oxidase amino acid
829	6	0.7	90	21	AA608931	Arabidopsis thalia
830	6	0.7	90	21	AA627162	Zea mays protein f
831	6	0.7	90	22	AA656826	Propionibacterium
832	6	0.7	90	22	AB629079	Peptide #1730 enco
833	6	0.7	90	22	AB638241	Peptide #5747 enco
834	6	0.7	90	22	AA620972	Human novel foetal
835	6	0.7	90	22	AA658867	Human brain expres
836	6	0.7	90	22	AA667421	Human bone marrow
837	6	0.7	90	22	AA690449	Human immune/haema
838	6	0.7	90	22	AA615249	Peptide #1683 enco
839	6	0.7	90	22	AA619033	Peptide #5467 enco
840	6	0.7	90	22	AA627710	Peptide #1747 enco
841	6	0.7	90	22	AA631674	Peptide #5711 enco
842	6	0.7	90	22	AA602993	Peptide #1675 enco
843	6	0.7	90	22	AA620585	Human secreted pro
844	6	0.7	90	23	AB677825	Amino acid sequenc
845	6	0.7	90	23	AB637044	Human peptide enco
846	6	0.7	91	19	AA660958	Streptococcus pneu
847	6	0.7	91	21	AA620176	Arabidopsis thalia
848	6	0.7	91	21	AA633052	Zea mays protein f
849	6	0.7	91	21	AA633055	Zea mays protein f
850	6	0.7	91	21	AA633059	Zea mays protein f
851	6	0.7	91	22	AB668643	Drosophila melanog
852	6	0.7	92	21	AA633250	Eucalyptus grandis
853	6	0.7	92	22	AA660374	Propionibacterium
854	6	0.7	92	22	AA631593	Novel human secret
855	6	0.7	92	23	AB631138	Human synthase-11k
856	6	0.7	93	22	AA683247	Human immune/haema
857	6	0.7	93	22	AA641804	Human polypeptide
858	6	0.7	93	23	ABP00429	Human OREX protein
859	6	0.7	93	23	ABP10159	Human OREX protein
860	6	0.7	94	23	AB680046	Mammary cancer 1 a
861	6	0.7	95	21	AA612453	Zea mays protein f
862	6	0.7	95	22	AA67445	Propionibacterium
863	6	0.7	95	22	AB609209	Novel human diagno
864	6	0.7	95	22	AA602377	Human polypeptide
865	6	0.7	95	23	ABP34436	Human glycoprotein
866	6	0.7	96	21	AA640831	Zea mays protein f
867	6	0.7	96	22	AA604390	Human polypeptide
868	6	0.7	96	23	ABP34208	Human OREX181 prot
869	6	0.7	96	23	ABP09306	Human OREX protein
870	6	0.7	97	21	AA645540	Human S100A10 prot
871	6	0.7	97	21	AA693605	Protein encoded by
872	6	0.7	97	22	AA651033	Propionibacterium
873	6	0.7	97	22	AB616109	Human nervous syst
874	6	0.7	97	22	AA650210	Human plasminogen-
875	6	0.7	97	22	AA671655	Human colon associ
876	6	0.7	97	23	ABP25789	Streptococcus poly
877	6	0.7	98	18	AA611296	S. pneumoniae SPOU
878	6	0.7	98	21	AA625335	Plus radiata cell
879	6	0.7	98	21	AA640830	Zea mays protein f
880	6	0.7	98	21	AA600891	Human secreted pro
881	6	0.7	98	21	AA659245	Tiliapia IGF-II E p
882	6	0.7	98	22	AA67006	Propionibacterium
883	6	0.7	98	22	AA655100	Propionibacterium
884	6	0.7	98	22	AA684210	Amino acid sequenc
885	6	0.7	98	22	AA673408	Human gene 14-enco
886	6	0.7	98	23	AB64233	Human albumin fusi
887	6	0.7	98	23	ABP05302	Human OREX protein
888	6	0.7	98	23	AB606791	Human sailor trans
889	6	0.7	99	20	AA635094	Chlamydia pneumoni
890	6	0.7	99	22	AA681895	S. epidermidis ope
891	6	0.7	99	23	ABP07507	Human OREX protein
892	6	0.7	100	21	AA649641	Arabidopsis thalia
893	6	0.7	100	21	AA661714	Arabidopsis thalia
894	6	0.7	100	22	AA644981	Propionibacterium
895	6	0.7	100	22	AB611918	Novel human diagno
896	6	0.7	101	20	AA634809	Chlamydia pneumoni
897	6	0.7	101	21	AA624035	Arabidopsis thalia
898	6	0.7	101	21	AA602231	Human secreted pro
899	6	0.7	101	22	AB669053	Drosophila melanog
900	6	0.7	101	22	AB635136	Peptide #2642 enco
901	6	0.7	101	22	AB620558	Protein #2557 enco
902	6	0.7	101	22	AA668332	Human bone marrow
903	6	0.7	101	22	AA628640	Peptide #2677 enco
904	6	0.7	101	22	AA603878	Peptide #2560 enco
905	6	0.7	101	22	AA692336	C glutamicum prote
906	6	0.7	101	23	AB637884	Human peptide enco
907	6	0.7	102	21	AA643467	Arabidopsis thalia
908	6	0.7	102	21	AA628001	Arabidopsis thalia
909	6	0.7	102	21	AA631521	Arabidopsis thalia
910	6	0.7	102	21	AA647020	Arabidopsis thalia
911	6	0.7	102	21	AA641892	Arabidopsis thalia
912	6	0.7	102	21	AB617720	Novel human diagno
913	6	0.7	102	22	AA693300	Human polypeptide
914	6	0.7	102	22	AA603006	Human polypeptide
915	6	0.7	102	22	AA607741	Human polypeptide
916	6	0.7	103	21	AA640839	Human OREX ORF603
917	6	0.7	103	21	AA641321	Human OREX ORF1085
918	6	0.7	103	21	AA607688	A human interleuk1
919	6	0.7	103	21	AA627367	Arabidopsis thalia
920	6	0.7	103	21	AA649640	Arabidopsis thalia
921	6	0.7	103	22	AA686927	Human immune/haema
922	6	0.7	103	22	AA660226	Pseudomonas mendoc
923	6	0.7	103	23	ABP43031	Human ovarian anti
924	6	0.7	103	23	ABP02144	Human OREX protein
925	6	0.7	103	23	ABP03730	Human OREX protein
926	6	0.7	104	20	AA629093	Human mammary asso
927	6	0.7	104	21	AA640961	Zea mays protein f
928	6	0.7	104	22	AA696567	Human reproductive
929	6	0.7	104	22	AA602347	Human polypeptide
930	6	0.7	104	23	ABP32685	Human ORE1658 prot
931	6	0.7	104	23	ABP05385	Human sailor trans
932	6	0.7	105	21	AB615396	E. coli mecJ mutan
933	6	0.7	105	21	AA635923	Zea mays protein f
934	6	0.7	105	21	AA600194	Human secreted pro
935	6	0.7	105	22	AA652438	Propionibacterium
936	6	0.7	105	22	AA600368	Human polypeptide
937	6	0.7	105	23	AA621778	Lung-specific amin
938	6	0.7	106	21	AA641441	Human OREX ORF1205
939	6	0.7	106	22	AA633516	Arabidopsis thalia
940	6	0.7	106	23	ABP33606	Human cancer relat
941	6	0.7	106	23	ABP33606	Human ORE2579 prot
942	6	0.7	106	23	AA647869	Mature placenta-de
943	6	0.7	107	13	AA622852	Promoter/secretion
944	6	0.7	107	14	AA647658	Promoter-secretion
945	6	0.7	107	17	AA699708	PKR11797-encoded p
946	6	0.7	107	21	AA618825	Zea mays protein f
947	6	0.7	107	21	AA658560	Arabidopsis thalia
948	6	0.7	107	22	AA640317	Propionibacterium
949	6	0.7	107	22	AB633413	Peptide #4064 enco
950	6	0.7	107	22	AB636627	Peptide #4133 enco
951	6	0.7	107	22	AA657390	Human brain expres
952	6	0.7	107	22	AA669781	Human bone marrow
953	6	0.7	107	22	AA617606	Peptide #4040 enco
954	6	0.7	107	22	AA635226	M tuberculosis R2
955	6	0.7	107	23	AB639413	Human peptide enco
956	6	0.7	107	23	AA648199	Human zinc finger
957	6	0.7	108	21	AA615029	Arabidopsis thalia
958	6	0.7	108	21	AA656859	Arabidopsis thalia
959	6	0.7	108	22	AB669222	Drosophila melanog

960	6	0.7	108	22	AAU53021	Propionibacterium
961	6	0.7	108	22	AAO10921	Human polypeptide
962	6	0.7	108	23	ABP04212	Human ORF protein
963	6	0.7	109	20	AAV13029	Human secreted pro
964	6	0.7	109	22	AAO01136	Human polypeptide
965	6	0.7	109	23	ABP03124	Human ORF protein
966	6	0.7	110	21	AAO31520	Arabidopsis thalia
967	6	0.7	110	21	AAO31520	Arabidopsis thalia
968	6	0.7	110	21	AAO31520	Arabidopsis thalia
969	6	0.7	110	21	AAO31520	Arabidopsis thalia
970	6	0.7	110	21	AAO31520	Arabidopsis thalia
971	6	0.7	110	21	AAO31520	Arabidopsis thalia
972	6	0.7	110	21	AAO31520	Arabidopsis thalia
973	6	0.7	110	21	AAO31520	Arabidopsis thalia
974	6	0.7	110	21	AAO31520	Arabidopsis thalia
975	6	0.7	110	21	AAO31520	Arabidopsis thalia
976	6	0.7	110	21	AAO31520	Arabidopsis thalia
977	6	0.7	110	21	AAO31520	Arabidopsis thalia
978	6	0.7	110	21	AAO31520	Arabidopsis thalia
979	6	0.7	110	21	AAO31520	Arabidopsis thalia
980	6	0.7	110	21	AAO31520	Arabidopsis thalia
981	6	0.7	110	21	AAO31520	Arabidopsis thalia
982	6	0.7	110	21	AAO31520	Arabidopsis thalia
983	6	0.7	110	21	AAO31520	Arabidopsis thalia
984	6	0.7	110	21	AAO31520	Arabidopsis thalia
985	6	0.7	110	21	AAO31520	Arabidopsis thalia
986	6	0.7	110	21	AAO31520	Arabidopsis thalia
987	6	0.7	110	21	AAO31520	Arabidopsis thalia
988	6	0.7	110	21	AAO31520	Arabidopsis thalia
989	6	0.7	110	21	AAO31520	Arabidopsis thalia
990	6	0.7	110	21	AAO31520	Arabidopsis thalia
991	6	0.7	110	21	AAO31520	Arabidopsis thalia
992	6	0.7	110	21	AAO31520	Arabidopsis thalia
993	6	0.7	110	21	AAO31520	Arabidopsis thalia
994	6	0.7	110	21	AAO31520	Arabidopsis thalia
995	6	0.7	110	21	AAO31520	Arabidopsis thalia
996	6	0.7	110	21	AAO31520	Arabidopsis thalia
997	6	0.7	110	21	AAO31520	Arabidopsis thalia
998	6	0.7	110	21	AAO31520	Arabidopsis thalia
999	6	0.7	110	21	AAO31520	Arabidopsis thalia
1000	6	0.7	110	21	AAO31520	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AA95987 standard: Protein: 919 AA.

AA95987;

05-DEC-2000 (first entry)

Moraxella catarrhalis BASB081 protein.

BASB081: Infection: otitis media; pneumonia; sinusitis; inflammation: therapy; antibacterial; antiinflammatory; vaccine;

Moraxella catarrhalis.

Key Location/Qualifiers

Peptide 1..30

Protein /label= Signal_peptide

FT /label= Mature_protein

FT MISC-difference 365

FT /note= "Ile in sequence of AA95988"

PN MO200052042-A1.

XX 08-SEP-2000.

Propionibacterium
Human polypeptide
Human ORF protein
Human secreted pro
Human polypeptide
Human ORF protein
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Human secreted pro
Propionibacterium
Propionibacterium
Propionibacterium
Staphylococcus epi
H. pylori ORF hpa
Secreted protein A
Human secreted pro
Human ORF ORF464
Human polypeptide
S. epidermidis ope
Human gene 4 encod
M. vaccae GroEL-ho
M. vaccae GroEL-ho
Anti-IL12 antibody
Anti-IL12 antibody
Human polypeptide
Human peptide #843
Peptide #875 encod
Protein #825 encod
Human brain expres
Human bone marrow
Peptide #853 encod
Human protein sequ
Peptide #869 encod
Peptide #828 encod
Human protein sequ
Human peptide encod
Human ORF protein
M vaccae GroEL hom
Human secreted pro
Human prostate can

Query Match 100.0%; Score 919; DB 21; Length 919;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 919; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKPVLPANRSEMPVLAATPLMTSOALAOQNNPANIINHVAHDTAIOAKAGNPVL 60
DB 1 MSKPVLPANRSEMPVLAATPLMTSOALAOQNNPANIINHVAHDTAIOAKAGNPVL 60

QY 61 LTPEDIOARLNAAGLNAPSOALDVNFDOSPTSRIGESQSPGLDMSVEETTPSL 120
DB 61 LTPEDIOARLNAAGLNAPSOALDVNFDOSPTSRIGESQSPGLDMSVEETTPSL 120

QY 121 EELFQESTEMKINPNDYIPYEQDQPNSEVVPPTLPKPKGLIKRLYARLNDGVNKKV 180
DB 121 EELFQESTEMKINPNDYIPYEQDQPNSEVVPPTLPKPKGLIKRLYARLNDGVNKKV 180

QY 181 PRLKAFYQSSQSGETSAIGSSHOKTEPYANIKALEDDTQESAMDNGSIPRLROFALV 240
DB 181 PRLKAFYQSSQSGETSAIGSSHOKTEPYANIKALEDDTQESAMDNGSIPRLROFALV 240

QY 241 AARAVGYDDIDLSIRNSIGEVYIHDLEPYIDYRAVEVGEADKAFITVADEVP 300
DB 241 AARAVGYDDIDLSIRNSIGEVYIHDLEPYIDYRAVEVGEADKAFITVADEVP 300

QY 301 LLIGDVFHGHKYEETKKNLTENASAEHGYFDGRWLDRSVVILPDMADVSLTYDGTQYR 360
DB 301 LLIGDVFHGHKYEETKKNLTENASAEHGYFDGRWLDRSVVILPDMADVSLTYDGTQYR 360

QY 361 FDEVEFTTIDPKTNOLTTDPDKLPVKRELLEQLLTVNNGEAVNLQAVRALSNDLIATRYF 420
DB 361 FDEVEFTTIDPKTNOLTTDPDKLPVKRELLEQLLTVNNGEAVNLQAVRALSNDLIATRYF 420

```
OY 421 NMVNTEIVPEREIONDOVSFEQSSSRTEPAQVDESTLEPIETVELTGMDSPI 480
DB 421 NMVNTEIVPEREIONDOVSFEQSSSRTEPAQVDESTLEPIETVELTGMDSPI 480
OY 481 EFSASNLIDDKLNLVAAKARHLTDMPPDRVLALNHDDGVNRSITLGRISDAVSARAILP 540
DB 481 EFSASNLIDDKLNLVAAKARHLTDMPPDRVLALNHDDGVNRSITLGRISDAVSARAILP 540
OY 541 DSESENEVIDLPERTALANRRTPADYVQSKKVPPLYVFVASDKPRDGOIGLGWSDTGRLV 600
DB 541 DSESENEVIDLPERTALANRRTPADYVQSKKVPPLYVFVASDKPRDGOIGLGWSDTGRLV 600
OY 601 TKFEHNLINRDGOQAQAEELRLSEDKKGVKLYAKRPLSHRINDOLRATLGOQEVFGHSTN 660
DB 601 TKFEHNLINRDGOQAQAEELRLSEDKKGVKLYAKRPLSHRINDOLRATLGOQEVFGHSTN 660
OY 661 GFDLSRTLEHEHSRSITLQNGWNRRTYSRLRYRLDKLKTQAPPETMODLPVDFVNGKFSQE 720
DB 661 GFDLSRTLEHEHSRSITLQNGWNRRTYSRLRYRLDKLKTQAPPETMODLPVDFVNGKFSQE 720
OY 721 ALLAGVAHKTYVADNLVNPMPRGYRQRYSLSEVGSSGLVSDANMAIARAGISGVYSEGDNAV 780
DB 721 ALLAGVAHKTYVADNLVNPMPRGYRQRYSLSEVGSSGLVSDANMAIARAGISGVYSEGDNAV 780
OY 781 GSNRAHQMTGGIOAGYIMSDNENHVPYRLREFAGGDSITGYAHDLSPTSDKGYLTGCG 840
DB 781 GSNRAHQMTGGIOAGYIMSDNENHVPYRLREFAGGDSITGYAHDLSPTSDKGYLTGCG 840
OY 841 VLAVGTAENYEFMKDLRLAVFGDIGNAYDKGFTNDTKIGAGVGRNASVGOVRVAVAT 900
DB 841 VLAVGTAENYEFMKDLRLAVFGDIGNAYDKGFTNDTKIGAGVGRNASVGOVRVAVAT 900
OY 901 GYKEEGNPILKHFICTPF 919
DB 901 GYKEEGNPILKHFICTPF 919

RESULT 2
AAY95988
ID AAY95988 standard; Protein; 889 AA.
XX
AC AAY95988;
DT 05-DEC-2000 (first entry)
XX
DE Moraxella catarrhalis BASB081 mature protein.
XX
KW BASB081; infection; otitis media; pneumonia; sinusitis;
KW inflammation; therapy; antibacterial; antinflammatory; vaccine;
KW diagnosis.
XX
OS Moraxella catarrhalis.
XX
FH Key location/Qualifiers
FT Misc-difference 335 /note="Val in sequence of AAY95987"
XX
PN WO200052042-A1.
XX
PD 08-SEP-2000.
XX
PF 23-FEB-2000; 2000MO-EP01468.
XX
PR 26-FEB-1999; 99GB-0004559.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Ruelle J;
XX
DR WPI; 2000-587296/55.
XX
DR N-PSDB; AAA50537.
XX
PT New BASB081 polypeptides from Moraxella catarrhalis and polynucleotides
```

```
PT encoding the polypeptides used for treating infections, or as a vaccine
PT for preventing infections, especially those caused by M. catarrhalis -
PS Claim 1; Page 61; 97pp; English.
XX
CC The present sequence is that of the Moraxella catarrhalis strain
CC Mc2391 (ATCC 43617) BASB081 mature protein. BASB081 is related by
CC amino acid sequence homology to Neisseria meningitidis omp5 outer
CC membrane protein. The sequence was deduced from PCR-amplified DNA
CC (see AAA50537), and shows 99.9% homology to the mature region of a
CC BASB081 gene translation product (see AAY95987). The invention
CC provides BASB081 polypeptides, polynucleotides, expression vectors,
CC host cells, and a process for producing a BASB081 polypeptide. Also
CC provided are vaccine compositions comprising a BASB081 polypeptide
CC or polynucleotide, and optionally at least 1 other M. catarrhalis
CC antigen. A method for diagnosing a M. catarrhalis infection
CC involves identifying a BASB081 polypeptide, or an antibody that is
CC immunospecific for it, in a sample. A therapeutic composition
CC useful in treating M. catarrhalis diseases in humans comprises an
CC antibody directed against a BASB081 polypeptide. The disease can
CC be a bacterial infection, e.g. otitis media in infants and children,
CC pneumonia in elderslies, sinusitis, nosocomial infections and
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed
CC speech learning, upper respiratory tract infection, and inflammation
CC of the middle ear.
XX
SQ Sequence 889 AA;
Query Match 85.7%; Score 788; DB 21; Length 889;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 888; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 31 QONNPANIIHVPAHDTAINDQAKAGNPVLTTPQIOARLNAAGINAKPOQALDVVNFD 90
DB 1 QONNPANIIHVPAHDTAINDQAKAGNPVLTTPQIOARLNAAGINAKPOQALDVVNFD 60
OY 91 DQSPISRIQEOSPPVLGIDMSYIEETTPLSLELFAQSTENGKINPNQIPIYQGEQPNSE 150
DB 61 DQSPISRIQEOSPPVLGIDMSYIEETTPLSLELFAQSTENGKINPNQIPIYQGEQPNSE 120
OY 151 VVVPPTLEPEKPGILKRLYARLFPNDGVNKPRLAKKPYQSSQSGTSAIGSSHOKTEPYA 210
DB 121 VVVPPTLEPEKPGILKRLYARLFPNDGVNKPRLAKKPYQSSQSGTSAIGSSHOKTEPYA 180
OY 211 NIKALEDITQESAMDINGSIPLRQATALVAARAVGYVDLSTIRNSIGEVYIINDLG 270
DB 181 NIKALEDITQESAMDINGSIPLRQATALVAARAVGYVDLSTIRNSIGEVYIINDLG 240
OY 271 EPVYIDYRAVEVRGEGADKAFTTVADVPLLIGDFVHHGKYEYFKKNIENASAEHGTFD 330
DB 241 EPVYIDYRAVEVRGEGADKAFTTVADVPLLIGDFVHHGKYEYFKKNIENASAEHGTFD 300
OY 331 GRWLDRSVVDVILPNTADVSLIYDTGYOYRDEVPVFTIOPKTNOLTPDPKLLVKKRELL 390
DB 301 GRWLDRSVVDVILPNTADVSLIYDTGYOYRDEVPVFTIOPKTNOLTPDPKLLVKKRELL 360
OY 391 EQLLTVMNGEAYNLQAAVRALSNDLIATRYFNMTETIYFPEREIONDOVSFEQSSSRT 450
DB 361 EQLLTVMNGEAYNLQAAVRALSNDLIATRYFNMTETIYFPEREIONDOVSFEQSSSRT 420
OY 451 EPAQVDESTLEPIETVELTGMDSPIEFSASNLIDDKLNLVAAKARHLTDMPPDRV 510
DB 421 EPAQVDESTLEPIETVELTGMDSPIEFSASNLIDDKLNLVAAKARHLTDMPPDRV 480
OY 511 LAIINHDDGVNRSITLGRISDAVSARAILPDESSENEVIDLPERTALANRRTPADYVQSKK 570
DB 481 LAIINHDDGVNRSITLGRISDAVSARAILPDESSENEVIDLPERTALANRRTPADYVQSKK 540
OY 571 VPLVVFVASDKPRDGOIGLGWSDTGRLVTKFEHNLINRDGOQAQAEELRLSEDKKGVKYL 630
DB 541 VPLVVFVASDKPRDGOIGLGWSDTGRLVTKFEHNLINRDGOQAQAEELRLSEDKKGVKYL 600
```

```

QY 631 YATKPLSHPLNDLRATIGYQOEYFGHSTNGFDLSTRLEHEISRSIIIONGGMNRTYSR 690
DB 601 YATKPLSHPLNDLRATIGYQOEYFGHSTNGFDLSTRLEHEISRSIIIONGGMNRTYSR 660
QY 691 YRLDKLKTQAPPEYMODLPVDFVNGKPSOEALLAGVAHVKTADLVNPMRCYRQYSLE 750
DB 661 YRLDKLKTQAPPEYMODLPVDFVNGKPSOEALLAGVAHVKTADLVNPMRCYRQYSLE 720
QY 751 VSSSGIVSANNATIRAGISGYISFGDNAYGSGNRAHQMTGTGIOAGYIMSDNFMHVPYRLR 810
DB 721 VSSSGIVSANNATIRAGISGYISFGDNAYGSGNRAHQMTGTGIOAGYIMSDNFMHVPYRLR 780
QY 811 FFAGGDOSIRGYAHDSLSIPSDKGYLTGGVLAAGTAETAYNEFKKDLRLAVFGDIGNAYD 870
DB 761 FFAGGDOSIRGYAHDSLSIPSDKGYLTGGVLAAGTAETAYNEFKKDLRLAVFGDIGNAYD 840
QY 871 KGFTMDTKIGAGVGRASPGVGVVDVANGVKEGPNRKLHFFIGRPF 919
DB 841 KGFTMDTKIGAGVGRASPGVGVVDVANGVKEGPNRKLHFFIGRPF 889

```

RESULT 3

```

AA95821
ID AAY95821 standard; Protein: 576 AA.
AC AAY95821;

```

```

DT 07-NOV-2000 (first entry)
XX

```

```

DE Haemophilus influenza nrl1 strain 289 BASB067 protein.
XX

```

```

KW BASB067; Outer membrane protein; antigen; vaccine; antidiotic;
antibacterial; screening; infection; diagnosis; therapy.
OS Haemophilus influenzae.
XX

```

```

FH Key
FT Location/Qualifiers
FT Peptide 1..20
FT Protein /label= Signal-Peptide
FT 21..576
FT /label= Mature-protein

```

```

PN MO200047737-A1.
XX

```

```

PD 17-AUG-2000.
XX

```

```

PF 04-FEB-2000; 2000MO-EP00887.
XX

```

```

PR 09-FEB-1999; 99GB-0002880.
XX

```

```

PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX

```

```

PI Ruelle J, Thonnard J;
XX

```

```

DR WPI: 2000-515059/46.
XX

```

```

N-PSDB; AAA50270.
XX

```

```

PS BASB067 polypeptide and polynucleotide from Haemophilus influenzae are
used for diagnosing and treating H. influenzae infections.
XX

```

```

Claim 1; Page 81-82; 87pp; English.
XX

```

```

CC The present sequence is that of outer membrane protein BASB067 of
CC Haemophilus influenzae non-typable (NHL) strain 289. BASB067 is
CC a surface expressed protein that is recognised by the immune system.
CC It shows homology to the protective surface antigen D15 of
CC H. influenzae and has a similar secondary structure. Its
CC N-terminal domain is predicted to contain a mixture of alpha-helix
CC and beta-strands, and could be used as a vaccine antigen. The
CC C-terminal domain is predicted to form a beta-barrel composed of
CC anti-parallel, amphipathic beta-strands. The external loops of
CC the beta-barrel of integral outer membrane proteins frequently
CC contain immunodominant B-cell epitopes, making the C-terminal

```

```

CC domain of BASB067 a strong candidate vaccine antigen. The
CC invention relates to recombinant materials and methods for the
CC production of BASB067 polypeptides and polynucleotides, for use
CC especially in therapeutic and prophylactic vaccines. It also
CC relates to methods for using such polypeptides and polynucleotides
CC in the prevention and treatment of microbial diseases, in diagnostic
CC assays for detecting diseases associated with microbial infections,
CC and assays for detecting expression or activity of BASB067
CC polypeptides or polynucleotides. Antibodies raised against
CC the mature portion of this BASB067 polypeptide can be used to treat
CC humans with H. influenzae disease.
XX

```

```

SQ Sequence 576 AA:

```

```

Query Match 1.2%; Score 11; DB 21; Length 576;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 882 GGVGRWASPVG 892
DB 537 GGVGRWASPVG 547

```

RESULT 4

```

AA95820
ID AAY95820 standard; Protein: 578 AA.
AC AAY95820;

```

```

DT 07-NOV-2000 (first entry)
XX

```

```

DE Haemophilus influenza strain Rd KW20 BASB067 protein.
XX

```

```

KW BASB067; Outer membrane protein; antigen; vaccine; antidiotic;
antibacterial; screening; infection; diagnosis; therapy.
OS Haemophilus influenzae.
XX

```

```

FH Key
FT Location/Qualifiers
FT Peptide 1..22
FT Protein /label= Signal-Peptide
FT 23..578
FT Domain /label= Mature-protein
FT 23..236
FT /note= "N-terminal domain"
FT Domain 237..578
FT /note= "C-terminal domain"

```

```

PN MO200047737-A1.
XX

```

```

PD 17-AUG-2000.
XX

```

```

PF 04-FEB-2000; 2000MO-EP00887.
XX

```

```

PR 09-FEB-1999; 99GB-0002880.
XX

```

```

PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX

```

```

PI Ruelle J, Thonnard J;
XX

```

```

DR WPI: 2000-515059/46.
XX

```

```

N-PSDB; AAA50269.
XX

```

```

PS BASB067 polypeptide and polynucleotide from Haemophilus influenzae are
used for diagnosing and treating H. influenzae infections.
XX

```

```

Claim 1; Page 78-80; 87pp; English.
XX

```

```

CC The present sequence is that of outer membrane protein BASB067 of
CC Haemophilus influenzae strain Rd KW20. BASB067 is a surface
CC expressed protein that is recognised by the immune system. It
CC shows 23% identity to the protective surface antigen D15 of
CC H. influenzae and has a similar secondary structure. Its

```


CC N-terminal domain is predicted to contain a mixture of alpha-helix
CC and beta-strands, and could be used as a vaccine antigen. The
CC C-terminal domain is predicted to form a beta-barrel composed of
CC anti-parallel, amphipathic beta-strands. The external loops of
CC the beta-barrels of integral outer membrane proteins frequently
CC contain immunodominant B cell epitopes, making the C-terminal
CC domain of BASB067 a strong candidate vaccine antigen. The
CC invention relates to recombinant materials and methods for the
CC production of BASB067 polypeptides and polynucleotides, for use
CC especially in therapeutic and prophylactic vaccines. It also
CC relates to methods for using such polypeptides and polynucleotides
CC in the prevention and treatment of microbial diseases, in diagnostic
CC assays for detecting diseases associated with microbial infections,
CC and assays for detecting expression or activity of BASB067
CC polypeptides or polynucleotides. Antibodies raised against
CC BASB067 can be used to treat humans with H. influenzae disease.
XX
SQ Sequence 578 AA;

Query Match 1.2%; Score 11; DB 21; Length 578;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 882 GVCGRASPVG 892
Db 539 GVCGRASPVG 549
|||||

RESULT 5
AA70579
ID AAY70579 standard; Protein; 262 AA.
XX
AC AAY70579;
XX
DT 04-JUL-2000 (first entry)
XX
DE Salmonella Pathogenicity Island 2 (SPI2) SseF protein.
XX
KW Salmonella Pathogenicity Island 2 locus; SPI2 locus; antigen carrier;
KW attenuation; gram-negative cell; vaccine; cytosolic; virulence; tumour;
KW anti-arteriosclerotic; anti-Alzheimer's; bactericide; hepatotropic;
KW anti-inflammatory; microbial infection; therapeutic; Salmonella infection;
KW Helicobacter pylori; stomach cancer; Herpes virus; Chlamydia pneumoniae;
KW Alzheimer's disease; arteriosclerosis; viral pathogen; Hepatitis virus;
KW cervical cancer; sseF; effector; type III secretion system.
XX
OS Salmonella typhimurium.
XX
PN WO200014240-A2.
XX
PD 16-MAR-2000.
XX
PE 03-SEP-1999; 99WO-EP06514.
XX
PR 04-SEP-1998; 98EP-0116827.
XX
PA (CREA-) CREATOGEN BIOSCIENCES GMBH.
XX
PI Hensel M, Guzman CA, Medina E, Apfel H, Hueck C;
XX WPI: 2000-256988/22.
XX DR N-PSDB: AA251991.
XX
PT Attenuated gram-negative Salmonella cells, comprising inactivated genes
PT in the SPI2 locus and useful for vaccinating against a range of
PT disorders associated with microbial infections such as stomach and
PT cervical cancers -
XX
PS Claim 18; Fig 23F; 180pp; English.
XX
CC The patent discloses attenuated gram-negative cells, especially
CC Salmonella, in which at least 1 gene in the Salmonella Pathogenicity
CC Island 2 (SPI2) locus has been inactivated resulting in attenuation/

CC reduction of virulence compared to the wild type cell. The attenuated
CC cells are used as carriers for presenting bacterial, viral or tumour
CC antigens to a host and are capable of expressing the antigen in a target
CC cell, especially a macrophage. The cells may therefore be used for the
CC preparation of a prophylactic or therapeutic composition for the
CC treatment of a chronic disease caused by a bacterium or virus, e.g.
CC Salmonella infection or a tumour. The cells may also be used to vaccinate
CC against a range of bacterial and viral pathogens e.g. Helicobacter pylori
CC (associated with stomach cancer), Chlamydia pneumoniae (associated with
CC arteriosclerosis and Alzheimer's disease), Borrelia burgdorferi,
CC Nanobacteria (found in the chronically diseased kidneys of patients
CC with crystalline deposits), Hepatitis virus (causative agent of
CC Hepatitis B and C and associated with liver cancer), Human papilloma
CC virus (HPV) (associated with cervical cancer) or Herpes virus.
CC The present sequence is the SseF protein, an effector of type III
CC secretion system, from the SPI2 locus of Salmonella. Inactivation of the
CC sse gene is useful for producing the attenuated cells.
XX
SQ Sequence 262 AA;

Query Match 1.0%; Score 9; DB 21; Length 262;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 PVLTPEQI 66
Db 43 PVLTPEQI 51
|||||

RESULT 6
AAR41333
ID AAR41333 standard; Protein; 851 AA.
XX
AC AAR41333;
XX
DT 22-APR-1994 (first entry)
XX
DE 113 KD ISGF-3alpha.
XX
KW 113 KD; 91 KD; 84 KD; ISGF-3alpha; interferon-related; receptor;
KW recognition factor; gene family; translation protein; tyrosine;
KW DNA binding protein; interferon-gamma; hairy cell leukaemia;
KW interferon therapy; chronic viral hepatitis; phosphorylation;
KW adjuvant therapy; tyrosine kinase.
XX
OS Homo sapiens.
XX
PN WO9319179-A.
XX
PD 30-SEP-1993.
XX
PE 19-MAR-1993; 93WO-US02569.
XX
PR 19-MAR-1992; 92US-0854296.
XX PR 23-NOV-1992; 92US-0980498.
XX
PA (UYRQ) UNIV ROCKEFELLER.
XX
PI Darnell JE, Fu X, Schindler CW, Shuai K;
XX WPI: 1993-320745/40.
XX DR N-PSDB: AA049164.
XX
PT Interferon receptor recognition factors - useful e.g. to treat
PT viral hepatitis, hairy cell leukaemia and to potentiate interferon
PT effects
XX
PS Claim 17; Fig 1; 131pp; English.
XX
CC The sequences given in AAR41333-35 represent the 113 KD, 91 KD and 84 KD
CC ISGF-3alpha proteins respectively. ISGF-3alpha is an interferon-
CC related receptor recognition factor which comprises several
CC substituents. The 113 KD, and the 91 and 84 KD proteins are derived

CC from two different but related genes. It is clear that a gene
 CC family exists and further members are likely to be found. The 91 kD
 CC protein has the capability of acting as a translation protein and as
 CC a DNA binding protein in response to interferon-gamma stimulation.
 CC These proteins participate in rapid phosphorylation and dephosphory-
 CC lation during the course of, and as part of their activity. This
 CC phosphorylation takes place in an interferon-dependant manner on
 CC specified tyrosine residues. These proteins may be used in
 CC conjunction with interferon therapy eg. to treat chronic viral
 CC hepatitis, hairy cell leukemia and for use with interferon in
 CC adjuvant therapy.

SO Sequence 851 AA:

Query Match 1.0%; Score 9; DB 14; Length 851;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ESTLEPVIE 465
 |||||||||
 DB 756 ESTLEPVIE 764

RESULT 7

ID AAR72077 standard; Protein: 851 AA.

AC AAR72077;

DT 27-SEP-1995 (first entry)

DE Recognition factor Stat13.

KW Signal transducer and activator of transcription; ISGF-3; STAT;

KW Stat13; receptor recognition factor; transcription factor;

KW cellular debilitation; derangement; dysfunction; interferon-alpha;

KW Interferon-gamma.

OS Homo sapiens.

PN WO9508629-A.

PD 30-MAR-1995.

PF 26-SEP-1994; 94WO-US10849.

PR 24-SEP-1993; 93US-0126588.

PR 24-SEP-1993; 93US-0126595.

PR 11-MAR-1994; 94US-0212184.

PR 11-MAR-1994; 94US-0212185.

PA (UYRQ) UNIV ROCKEFELLER.

PI Darnell JE, Schindler CW, Shuai K, Wen Z, Zhong Z;

DR WPI: 1995-139598/18.

DR N-PSDB; AA089335.

PT Receptor recognition factor implicated in transcriptional

PT stimulation of genes useful in drug screening assays and/or

PT for treating cellular debilitations, derangements and/or

PT dysfunctions, etc.

XX Disclosure: Page 78-81; 160pp; English.

XX The sequences of cDNA encoding receptor recognition factors having

CC mol. wt. of 113 kDa (Stat13), 91 kDa (Stat92) and 84 kDa (Stat84)

CC are given in AA089335-37 and the deduced amino acid sequences of the

CC STAT proteins in AAR72077-79. These ISGF-3-derived proteins are

CC activated by binding of interferon-alpha (all 3 Stat proteins) or

CC Interferon-gamma (Stat91) to cell receptors.

XX Sequence 851 AA:

Query Match 1.0%; Score 9; DB 16; Length 851;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ESTLEPVIE 465
 |||||||||
 DB 756 ESTLEPVIE 764

RESULT 8

ID AAW03166 standard; Protein: 851 AA.

AC AAW03166;

DT 24-OCT-1996 (first entry)

DE Human STAT2.

KW STAT2; signal transducer and activator of transcription;

KW DNA binding protein; ligand; receptor; oncogenes; inflammation;

KW autoimmune disease; antagonist; therapy; STAT13.

OS Homo sapiens.

PN Key Location/Qualifiers

FT Domain 396..506

FT /label= "DNA binding domain

FT /note= "Claim 3, page 111"

PN WO9620954-A2.

PD 11-JUL-1996.

PF 28-DEC-1995; 95WO-US17025.

PR 06-JAN-1995; 95US-0369796.

PA (UYRQ) UNIV ROCKEFELLER.

PI Darnell JE, Horvath CM, Wen Z, Zhong Z;

DR WPI: 1996-33941/33.

DR N-PSDB; AAT31275.

PT New STAT protein DNA-binding domain peptide(s) - useful for

PT diagnosing, preventing or treating cellular dysfunction, e.g.

PT oncogenesis, inflammation, parasitic disease or autoimmunity

XX Disclosure: Page 63-66; 138pp; English.

XX Signal transducer and activator of transcription (STAT) protein STAT2

CC (AAW03166), also known as STAT113, is a 113 kDa protein having a dual

CC purpose, i.e. signal transduction from ligand-activated receptor

CC kinase complexes followed by nuclear translocation and DNA binding

CC to activate transcription. The amino acid sequence of STAT2 was

CC deduced from a cDNA clone (AAT31275) derived from HeLa cells. STAT2

CC includes a DNA-binding domain (see also AAW03175) capable of both

CC receptor recognition and message delivery via DNA binding in a

CC receptor-ligand specific manner. STAT proteins and their DNA

CC binding domains (see also AAW03165, AAW03167-76) are useful for

CC screening antagonists used to inhibit STAT-mediated signal

CC transduction and activation of transcription.

XX Sequence 851 AA:

SO Query Match 1.0%; Score 9; DB 17; Length 851;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ESTLEPVIE 465
 |||||||||

XX Sequence 851 AA;
SQ
Query Match 1.0%; Score 9; DB 22; Length 851;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 457 ESTLEPVIE 465
DB 756 ESTLEPVIE 764
RESULT 11
AAE14649
ID AAE14649 standard; Protein; 851 AA.
XX
AC AAE14649;
XX
DT 16-JUL-2002 (first entry)
XX
DE Human STAT2 protein.
XX
KW Signal transducer and activator of transcription; STAT2;
KW drug development; drug discovery; crystal; inflammation; allergy;
KW asthma; leukaemia; anaemia; neutropenia; thrombocytopaenia;
KW cancer; obesity; viral disease; growth retardation; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 1..130
FT /note= "Conserved N-terminal domain of the STAT family"
XX
PN US6312887-B1.
XX
PD 06-NOV-2001.
XX
PE 24-APR-2000; 2000US-0556273.
XX
PR 23-JAN-1998; 98US-0012710.
XX
XX
PA (UVRQ) UNIV ROCKEFELLER.
XX
PI Vinkemeier U, Moarefi I, Darnell JE, Kuriyan J;
XX
DR WPI: 2002-033337/04.
XX
PT Identifying compounds that bind to signal transducer and activator of
PT transcription proteins, useful for the production of new drugs
PS Example; Column 35-40; 44pp: English.
XX
XX The invention relates to methods for detecting compounds that bind to
CC signal transducer and activator of transcription (STAT) proteins for
CC the discovery and development of new drug compounds based on the
CC structural properties of the protein crystal. The methods include:
CC identifying a compound that binds to the N-terminal domain of a
CC STAT protein, identifying a compound that enhances or diminishes the
CC binding of the dimeric STAT proteins to each other and/or their nucleic
CC acid binding site; or identifying a compound that enhances or diminishes
CC the ability of STAT protein dimers to induce the expression
CC of a gene operably under the control of a promoter containing at least
CC two adjacent weak binding sites for STAT protein dimers. The methods
CC are used for identifying new drugs. An antagonist of STAT N-terminal
CC dimeric interactions that inhibits the binding of the STAT dimers to
CC adjacent weak binding sites on a promoter of a gene, could be useful
CC as drugs in the treatment of diseases, e.g. inflammation, allergy,
CC asthma and leukaemia. On the other hand, an agonist of N-terminal
CC dimeric interactions between STAT dimers, can be used as drugs in the
CC treatment of diseases e.g. anaemia, neutropenia, thrombocytopaenia,
CC cancer, obesity, viral diseases and growth retardation.
XX The present sequence is human STAT2 protein.

SQ Sequence 851 AA;
SQ
Query Match 1.0%; Score 9; DB 23; Length 851;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 457 ESTLEPVIE 465
DB 756 ESTLEPVIE 764
RESULT 12
AAE15173
ID AAE15173 standard; Protein; 851 AA.
XX
AC AAE15173;
XX
DT 07-MAR-2002 (first entry)
XX
DE Human Stat2 (signal transducer and activator of transcription) protein.
XX
KW Human; SOCS; suppressor of cytokine signalling; autoimmune disorder; Jak;
KW Janus kinase; signal transducer and activator of transcription; Stat;
KW transplant acceptance; systemic lupus erythematosus; glomerulonephritis;
KW rheumatoid arthritis; Wegener's granulomatosis; chronic active hepatitis;
KW atopic dermatitis; multiple sclerosis; myasthenia gravis; vasculitis;
KW Crohn's disease; haemolytic anaemia; nephrotic syndrome; dermatological;
KW diabetes mellitus; thyroiditis; inflammatory bowel disease; nephrotropic;
KW immunosuppressive; antiinflammatory; antirheumatic; antiarthritic;
KW hepatotropic.
XX
XX Homo sapiens.
XX
OS WO200179555-A2.
XX
PN 25-OCT-2001.
XX
PD 13-APR-2001; 2001WO-US12131.
XX
PE 14-APR-2000; 2000US-0549654.
XX
PR (MILL-) MILLENNIUM PHARM INC.
XX
XX
PA Hancock WW, Ozkaynak E;
XX
PI
XX
DR WPI: 2002-034368/04.
XX
DR N-PSDB; AAD24305.
XX
PT Monitoring transplant acceptance or autoimmune disease, useful e.g. for
PT assessing therapy, comprises measuring levels of Stat or their
PT inhibitors
XX
XX Example; Fig 2: 218pp: English.
XX
XX The present invention relates to a method for monitoring acceptance of a
CC transplant or an autoimmune disease in a mammal. The method comprising
CC determining the amount of at least one of Stat4 (signal transducer and
CC activator of transcription), Stat6, SOCS1 (suppressor of cytokine
CC signalling) or SOCS3 mRNA or protein in a sample of the transplant, taken
CC from the host or an affected tissue sample. Stats are activated by
CC receptor-associated Janus kinases (Jaks) which include Jak1, Jak2, Tyk2,
CC Jak3. The method is used to determine whether acceptance of a transplant
CC has been induced or to determine if autoimmune disorders (systemic
CC lupus erythematosus, glomerulonephritis, rheumatoid arthritis, Wegener's
CC granulomatosis, chronic active hepatitis, atopic dermatitis, multiple
CC sclerosis, myasthenia gravis, haemolytic anaemia, nephrotic syndrome,
CC thyroiditis, diabetes mellitus, Crohn's disease, inflammatory bowel
CC disease and vasculitis) are being treated successfully and may be used
CC to adjust treatment regimes. The present sequence is human Stat2 protein.
XX
SQ Sequence 851 AA;
SQ
Query Match 1.0%; Score 9; DB 23; Length 851;

Best Local Similarity 100.0%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465
| | | | | | | |
Db 756 ESTLEPVIE 764

RESULT 13
ABG25555
ID ABG25555 standard; Protein: 114 AA.

AC ABG25555;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #25546.

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB: AAS89742.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID No 55914; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 114 AA;

Query Match 0.9%; Score 8; DB 22; Length 114;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 809 LRFAGCD 816
| | | | | | | |
Db 52 LRFAGCD 59

RESULT 14
ABB47876
ID ABB47876 standard; Protein: 331 AA.

AC ABB47876;

DT 05-FEB-2002 (first entry)

DE Listeria monocytogenes protein #580.

KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.

OS Listeria monocytogenes.

PN WO200177335-A2.

PD 18-OCT-2001.

PF 11-APR-2001; 2001WO-FR01118.

PR 11-APR-2000; 2000FR-0004629.

XX (INSP) INST PASTEUR.

PI Buchrieser C, Frangoul L, Couve E, Rusnlok C, Eshti H, Dehoux P;

PI Dussurget O, Chetouani F, Medjari H, Glaser P, Kunst F, Cossart P;

PI Daniels J, Goebel W, Kreft J, Kunz M, Ng E, Vazquez-Boland JA;

PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;

PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;

PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

PI Meduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;

PI Rose M, Voss H;

DR WPI: 2002-010914/01.

PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides -

PS Claim 6; SEQ ID No 581; 192pp; French.

XX The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccine compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 331 AA;

Query Match 0.9%; Score 8; DB 23; Length 331;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 839 GOVLAVGT 846
Db 38 GOVLAVGT 45

RESULT 15

ABG25558
ID ABG25558 standard; Protein; 371 AA.
AC
XX ABG25558;
AC

DT 18-FEB-2002 (first entry)
XX
XX

DE Novel human diagnostic protein #25549.
XX
XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
OS

PN WO200175067-A2.
XX

PD 11-OCT-2001.
XX

PE 30-MAR-2001: 2001MO-US08631.
XX

PR 31-MAR-2000: 2000US-0540217.
XX 23-AUG-2000: 2000US-0649167.
XX

PA (HYSE-) HYSEQ INC.
XX

PI Drmanac RT, Liu C, Tang YT;
XX

DR WPI: 2001-639362/73.
XX N-PSDB: AAS89745.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX

PS Claim 20; SEQ ID NO 55917; 103pp; English.
XX

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 371 AA;

Query Match 0.9%; Score 8; DB 22; Length 371;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 809 LRFAGCD 816
Db 112 LRFAGCD 119

Db 185 LRFAGCD 192

RESULT 16

ABG17710
ID ABG17710 standard; Protein; 474 AA.
AC
XX ABG17710;
AC

DT 18-FEB-2002 (first entry)
XX
XX

DE Novel human diagnostic protein #17701.
XX
XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
OS

PN WO200175067-A2.
XX

PD 11-OCT-2001.
XX

PE 30-MAR-2001: 2001MO-US08631.
XX

PR 31-MAR-2000: 2000US-0540217.
XX 23-AUG-2000: 2000US-0649167.
XX

PA (HYSE-) HYSEQ INC.
XX

PI Drmanac RT, Liu C, Tang YT;
XX

DR WPI: 2001-639362/73.
XX N-PSDB: AAS81897.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX

PS Claim 20; SEQ ID NO 48069; 103pp; English.
XX

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 474 AA;

Query Match 0.9%; Score 8; DB 22; Length 474;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 809 LRFAGCD 816
Db 112 LRFAGCD 119

```

RESULT 17
AAW56318
ID AAW56318 standard; Protein; 582 AA.
XX
AC AAW56318;
XX
DT 27-JUL-1998 (first entry)
XX
DE Infectious laryngotracheitis virus antigenic protein.
XX
KM Antigenic protein; Infectious laryngotracheitis virus; ILTV;
KM recombinant virus; fowlpox; avipox; chickenpox; vaccine; chicken;
KM pheasant; turkey.
XX
OS Infectious laryngotracheitis virus.
XX
PN WO9807866-A1.
XX
PD 26-FEB-1998.
XX
PF 21-AUG-1997; 97WO-JP02912.
XX
PR 21-AUG-1996; 96JP-0238580.
XX
PA (JAPG ) NIPPON ZEON KK.
XX
PI Okuda T, Tsuzaki Y;
XX
DR WPI: 1998-169170/15.
DR N-PSDB; AAW22805.
XX
PT Antigenic protein from infectious laryngotracheitis virus and DNA
PT coding for it - which is incorporated into recombinant avian virus
PT for vaccine production
XX
PS Claim 1; Pages 38-41; 69pp; Japanese.
XX
CC The present sequence represents an antigenic protein originating in
CC infectious laryngotracheitis virus (ILTV). The DNA sequence can be
CC incorporated into recombinant viruses (such as fowlpox, avipox or
CC chickenpox virus) and used to produce vaccines. The vaccines are an
CC effective means of protecting birds such as chickens, pheasants or
CC turkeys against infectious laryngotracheitis.
XX
SQ Sequence 582 AA;

Query Match 0.9%; Score 8; DB 19; Length 582;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 213 KAALEDIT 220
DB 277 KAALEDIT 284

RESULT 18
ABB62368
ID ABB62368 standard; Protein; 1373 AA.
XX
AC ABB62368;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13896.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX

```

```

PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
DR N-PSDB; ABL06471.
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 13896; 21pp + Sequence Listing; English.
XX
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1373 AA;

Query Match 0.9%; Score 8; DB 22; Length 1373;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LPEEIOIA 68
DB 913 LPEEIOIA 920

RESULT 19
AAR44425
ID AAR44425 standard; Protein; 29 AA.
XX
XX
AC AAR44425;
XX
DT 26-MAY-1994 (first entry)
XX
DE Mutant alpha-amylase signal peptide.
XX
KM Bacillus licheniformis; alpha-amylase; signal peptide;
KM peptidase; mature protein; recognition site; mutant;
KM cleavage site; interleukin; IL-3.
XX
OS Bacillus licheniformis.
XX
XX
PN EP572088-A.
XX
PD 01-DEC-1993.
XX
PF 25-MAY-1993; 93EP-0201500.
XX
PR 25-MAY-1992; 92EP-0201492.
XX
PA (KONN ) GIST-BROCADES NV.
XX
PI Bonekamp AJ, Van Tilborg MWEW;
XX
DR WPI: 1993-379003/48.
XX

```

PT New mutant signal sequences with fewer potential cleavage sites -
PT reduce processing ambiguity of signal peptidase(s)
PS Claim 7; Page 15; 23pp; English.
XX
XX
CC The two cleavage sites of B. licheniformis alpha-amylase signal
CC peptide may give rise to two different mature proteins (AAR44429).
CC To obtain a proper mature prod., the second recognition site
CC has to be destroyed without interfering with the first recognition
CC site. The combination of the B. licheniformis alpha-amylase
CC signal sequence with human IL-3 shows a third possible cleavage
CC site since mature IL-3 starts with Ala (AAR51699). In practice
CC the third site is not used, as the signal peptidase is not
CC active when the Ala-X-Ala is upstream of a pro.
CC Mutant signal sequences with fewer potential cleavage sites
CC reduce the processing ambiguity of signal peptidases.
CC Mutated Bacillus alpha-amylase sequences in combination with
CC IL-3 are illustrated in AAR44425-28.
CC The oligonucleotide probes used to obtain the desired mutations
CC are given in AAO52516-25.
XX
SQ Sequence 29 AA:
Query Match 0.88; Score 7; DB 14; Length 29;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 166 KRLYARL 172
|||||||
Db 5 KRLYARL 11
RESULT 20
AAR44426
ID AAR44426 standard; Protein: 29 AA.
XX
AC AAR44426;
XX
DT 26-MAY-1994 (first entry)
XX
DE Mutant alpha-amylase signal peptide.
XX
KW Bacillus licheniformis; alpha-amylase; signal peptide;
KW peptidase; mature protein; recognition site; mutant;
KW cleavage site; interleukin; IL-3.
XX
OS Bacillus licheniformis.
XX
PN EP572088-A.
PD 01-DEC-1993.
XX
PF 25-MAY-1993; 93EP-0201500.
XX
PR 25-MAY-1992; 92EP-0201492.
XX
PA (KONN) GIST-BROCADES NV.
PI Bonekamp AJ, Van Tilborg MWEM;
XX
DR WPT; 1993-379003/48.
XX
PT New mutant signal sequences with fewer potential cleavage sites -
PT reduce processing ambiguity of signal peptidase(s)
PS Claim 7; Page 15; 23pp; English.
XX
XX
CC The two cleavage sites of B. licheniformis alpha-amylase signal
CC peptide may give rise to two different mature proteins (AAR44429).
CC To obtain a proper mature prod., the second recognition site
CC has to be destroyed without interfering with the first recognition
CC site. The combination of the B. licheniformis alpha-amylase
CC signal sequence with human IL-3 shows a third possible cleavage

CC site since mature IL-3 starts with Ala (AAR51699). In practice
CC the third site is not used, as the signal peptidase is not
CC active when the Ala-X-Ala is upstream of a pro.
CC Mutant signal sequences with fewer potential cleavage sites
CC reduce the processing ambiguity of signal peptidases.
CC Mutated Bacillus alpha-amylase sequences in combination with
CC IL-3 are illustrated in AAR44425-28.
CC The oligonucleotide probes used to obtain the desired mutations
CC are given in AAO52516-25.
XX
SQ Sequence 29 AA:
Query Match 0.88; Score 7; DB 14; Length 29;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 166 KRLYARL 172
|||||||
Db 5 KRLYARL 11
RESULT 21
AAR44427
ID AAR44427 standard; Protein: 29 AA.
XX
AC AAR44427;
XX
DT 26-MAY-1994 (first entry)
XX
DE Mutant alpha-amylase signal peptide.
XX
KW Bacillus licheniformis; alpha-amylase; signal peptide;
KW peptidase; mature protein; recognition site; mutant;
KW cleavage site; interleukin; IL-3.
XX
OS Bacillus licheniformis.
XX
PN EP572088-A.
PD 01-DEC-1993.
XX
PF 25-MAY-1993; 93EP-0201500.
XX
PR 25-MAY-1992; 92EP-0201492.
XX
PA (KONN) GIST-BROCADES NV.
PI Bonekamp AJ, Van Tilborg MWEM;
XX
DR WPT; 1993-379003/48.
XX
PT New mutant signal sequences with fewer potential cleavage sites -
PT reduce processing ambiguity of signal peptidase(s)
PS Claim 7; Page 15; 23pp; English.
XX
XX
CC The two cleavage sites of B. licheniformis alpha-amylase signal
CC peptide may give rise to two different mature proteins (AAR44429).
CC To obtain a proper mature prod., the second recognition site
CC has to be destroyed without interfering with the first recognition
CC site. The combination of the B. licheniformis alpha-amylase
CC signal sequence with human IL-3 shows a third possible cleavage
CC site since mature IL-3 starts with Ala (AAR51699). In practice
CC the third site is not used, as the signal peptidase is not
CC active when the Ala-X-Ala is upstream of a pro.
CC Mutant signal sequences with fewer potential cleavage sites
CC reduce the processing ambiguity of signal peptidases.
CC Mutated Bacillus alpha-amylase sequences in combination with
CC IL-3 are illustrated in AAR44425-28.
CC The oligonucleotide probes used to obtain the desired mutations
CC are given in AAO52516-25.
XX
SQ Sequence 29 AA:

Query Match 0.8%; Score 7; DB 14; Length 29;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 KRUYARL 172
 |||||
 DB 5 KRLUYARL 11

RESULT 22

AAAR44428
 ID AAR44428 standard; Protein; 29 AA.

AC AAR44428;

DT 26-MAY-1994 (first entry)

DE Mutant alpha-amylase signal peptide.

KW Bacillus licheniformis; alpha-amylase; signal peptide;

KW peptidase; mature protein; recognition site; mutant;

KW cleavage site; interleukin; IL-3.

OS Bacillus licheniformis.

PN EP572088-A.

PD 01-DEC-1993.

PF 25-MAY-1993; 93EP-0201500.

PR 25-MAY-1992; 92EP-0201492.

PA (KONN) GIST-BROCADES NV.

PI Bonekamp AJ, Van Tilborg MMEW;

DR WPI; 1993-379003/48.

PT New mutant signal sequences with fewer potential cleavage sites -

PS reduce processing ambiguity of signal peptidase(s)

XX Claim 7; Page 15; 23pp; English.

CC The two cleavage sites of B. licheniformis alpha-amylase signal

CC peptide may give rise to two different mature proteins (AAR44429).

CC To obtain a proper mature prod., the second recognition site

CC has to be destroyed without interfering with the first recognition

CC site. The combination of the B. licheniformis alpha-amylase

CC signal sequence with human IL-3 shows a third possible cleavage

CC site since mature IL-3 starts with Ala (AAR51699). In practice

CC the third site is not used, as the signal peptidase is not

CC active when the Ala-X-Ala is upstream of a Pro.

CC Mutant signal sequences with fewer potential cleavage sites

CC reduce the processing ambiguity of signal peptidases.

CC Mutated Bacillus alpha-amylase sequences in combination with

CC IL-3 are illustrated in AAR44425-28.

CC The oligonucleotide probes used to obtain the desired mutations

CC are given in AAR52516-25.

XX Sequence 29 AA;

Query Match 0.8%; Score 7; DB 14; Length 29;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 KRLUYARL 172
 |||||

DB 5 KRLUYARL 11

RESULT 23

AAAR74559
 ID AAR74559 standard; Peptide; 34 AA.

AC AAR74559;

DT 04-JAN-1996 (first entry)

DE Protease Nexin-1 N-terminal peptide.

KW Protease Nexin-1 N-terminal peptide; PN-1; variants;

KW protease inhibitors; urokinase inhibitor; inflammation;

KW tumour cell invasion; elastase inhibitor; emphysema;

KW septic shock; wound healing.

OS Homo sapiens.

PN WO9511987-A1.

PD 04-MAY-1995.

PF 28-OCT-1994; 94WO-US11624.

PR 29-OCT-1993; 93US-0144758.

PA (INCY-) INCYTE PHARM INC.

PI Braxton SM, Scott RW;

DR WPI; 1995-178879/23.

PT New protease nexin-1 variant with altered active site - for treating

PT inflammation etc.; also proteins modified by polyethylene glycol

PT attached to Cys thiol residues, related DNA and nexin fusion

PS proteins.

XX Example A; Page 73; 122pp; English.

CC AAR74559 is the protease Nexin-1 (PN-1) N-terminal peptide. It was used

CC to isolate the complete PN-1 sequence, from which the active site

CC variants described in AAR74553-R74546 were derived. These variants have

CC different protease specificities and/or increased rate association

CC constants compared to the native PN-1. The variants can be used as

CC urokinase inhibitors (to treat inflammation and tumour cell invasion),

CC as elastase inhibitors (to treat emphysema and septic shock) and as

CC wound healing stimulators.

XX Sequence 34 AA;

Query Match 0.8%; Score 7; DB 16; Length 34;

Best Local Similarity 100.0%; Pred. No. 61;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 PLSLLEL 123
 |||||

DB 5 PLSLLEL 11

RESULT 24

AAAW51936
 ID AAW51936 standard; Peptide; 34 AA.

AC AAW51936;

DT 26-OCT-1998 (first entry)

DE Human protease nexin I N-terminal peptide.

KW Nexin I type alpha; PN-1; serine protease inhibitor; serpin; human;

KW antinflammatory; cysteine-PEGylated protein; polyethylene glycol;

XX PEG.

OS Homo sapiens.

PR 28-MAR-1986; 86US-0845864.
PR 29-MAR-1985; 85US-0717321.
XX
XX (BIOT-) BIOTECHNICA INT INC.
PA
XX Stephens MA, Rudolph CF, Hannett NM, Stassi DL, Pero JG;
XX WPI; 1986-278825/42.
DR N-PSDB; AAN60673.
XX
PT Vectors useful for transforming Gram positive bacteria - contg.
PT secretory signal encoding sequence of Bacillus licheniformis
PT alpha-amylase gene
XX
PS Example; Fig 8; 60pp; English.
XX
CC The vectors of the invention may be used to transform Gram positive
CC bacteria for prodn. of polypeptides e.g. growth hormone. Using the
CC vectors contg. a gene for alkaline phosphatase fused to a gene
CC encoding growth hormone, expression can be monitored by observing
CC enzymatic activity.
CC
SQ Sequence 41 AA;

Query Match 0.8%; Score 7; DB 7; Length 41;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 KRLYARL 172
|||
Db 5 KRLYARL 11

RESULT 27
AAB65038
ID AAB65038 standard; peptide: 41 AA.
XX
AC AAB65038;
XX
DT 23-MAR-2001 (first entry)
XX
DE Gene #5 associated peptide #6.
XX
KM Secreted protein; gene therapy; vaccine; cancer; leukemia;
KM autoimmune disease; allergy; inflammation; graft rejection;
KM hyperproliferation; cardiovascular; infection.
XX
OS Homo sapiens.
XX
XX WO200075375-A1.
PN
XX 14-DEC-2000.
PD
XX 02-JUN-2000; 2000WO-US15187.
PF
XX 07-JUN-1999; 99US-0137725.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Birse CE, Duan RD, Soppet DR, Rosen CA, Shi Y;
PI Lafleur DW, Olsen HS, Edner R, Florence KA, Nl J, Young PE;
XX WPI; 2001-061741/07.
DR
XX
PT Nucleic acids encoding 26 human secreted polypeptides, useful e.g. for
PT preventing, diagnosing and/or treating cancers and for promoting wound
PT healing -
XX
PS Disclosure; Page 23; 530pp; English.
XX
CC The present invention relates to 26 secreted human proteins. The
CC proteins may be used in the prevention, diagnosis and treatment of
CC diseases associated with inappropriate polypeptide expression.

CC For example, they may be used in gene therapy or in vaccines.
CC Typical of diseases which are potentially treatable are cancers
CC (including leukemia), autoimmune diseases, allergies, inflammation,
CC graft rejection, hyperproliferation, cardiovascular diseases
CC (particularly critical limb ischemia and coronary disease) and any
CC involving abnormal angiogenesis, neurodegeneration and/or
CC infectious diseases.
XX
SQ Sequence 41 AA;

Query Match 0.8%; Score 7; DB 22; Length 41;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 DRVLAIN 514
|||
Db 4 DRVLAIN 10

RESULT 28
AAP60715
ID AAP60715 standard; Protein; 42 AA.
XX
AC AAP60715;
XX
DT 23-JUL-1991 (first entry)
XX
DE Sequence of the fusion of the alpha-amylase signal sequence to
DE alkaline phosphatase in the vector p2/38.
XX
XX Vectro: Gram positive bacteria.
XX
XX Bacillus licheniformis.
XX
FH Key Location/Qualifiers
FT Peptide 1..29
FT Region /label= signal
FT /label= linker
FT Protein 40..42
FT /note= "starts at AA residue 12"
XX
XX WO8605812-A.
PN
XX 09-OCT-1986.
PD
XX 28-MAR-1986; 86WO-US00636.
PF
XX 28-MAR-1986; 86US-0845864.
PR 29-MAR-1985; 85US-0717321.
XX
PA (BIOT-) BIOTECHNICA INT INC.
XX
PI Stephens MA, Rudolph CF, Hannett NM, Stassi DL, Pero JG;
XX WPI; 1986-278825/42.
DR N-PSDB; AAN60674.
XX
PT Vectors useful for transforming Gram positive bacteria - contg.
PT secretory signal encoding sequence of Bacillus licheniformis
PT alpha-amylase gene
XX
PS Example; Fig 10; 60pp; English.
XX
CC The vectors of the invention may be used to transform Gram positive
CC bacteria for prodn. of polypeptides e.g. growth hormone. Using the
CC vectors contg. a gene for alkaline phosphatase fused to a gene
CC encoding growth hormone, expression can be monitored by observing
CC enzymatic activity.
XX
SQ Sequence 42 AA;

Query Match 0.8%; Score 7; DB 7; Length 42;

Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 KRLYARL 172
|||||||
DB 5 KRLYARL 11

RESULT 29

AAV64841
ID AAV64841 standard; Protein; 51 AA.

AC AAV64841;

DT 01-FEB-2000 (first entry)

DE Human 5' EST related polypeptide SEQ ID NO:1002.

XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification.

XX Homo sapiens.

PN MO9953051-A2.

PD 21-OCT-1999.

PF 09-APR-1999; 99WO-1B00712.

XX 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-038446/03.

XX N-PSDB; AA242455.

PT Novel secreted protein 5' expressed sequence tag sequences used in
diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS Claim 3; Page 657-658; 837pp; English.

CC AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
sequences, corresponding to human secreted proteins. AAV64651 to
CC AAV65438 represent the EST-related proteins corresponding to AA242265 to
CC AA243052. The 5' ESTs can be used for producing secreted human gene
products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
location, development stage, rate, and quantity of protein synthesis, as
well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value and the identification of new secreted proteins is
CC valuable. AA242249 to AA242264 and AAV64644 to AAV64650 represent
sequences used in the exemplification of the present invention.

XX Sequence 51 AA;

Query Match 0.8%; Score 7; DB 21; Length 51;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 LTDGILM 475
|||||||
DB 2 LTDGILM 8

RESULT 30

AAU22430
ID AAU22430 standard; Protein; 51 AA.

AC AAU22430;

DT 18-DEC-2001 (first entry)

DE Human cardiovascular system antigen polypeptide SEQ ID No 1204.

XX Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW Chicken; Sheep; Immunosuppressive; antiarthritic; vasotropic; dog;
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; noctropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; tissue regeneration;
KW anti-infertility.

XX Homo sapiens.

PN WO20015321-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01340.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

XX 14-JUL-2000; 2000US-0217496.

PR 26-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

XX 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

XX 22-AUG-2000; 2000US-0226868.

PR 23-AUG-2000; 2000US-0227182.

XX 30-AUG-2000; 2000US-0227009.

PR 01-SEP-2000; 2000US-0228924.

XX 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000: 2000US-0229344.
PR 01-SEP-2000: 2000US-0229345.
PR 05-SEP-2000: 2000US-0229509.
PR 05-SEP-2000: 2000US-0229513.
PR 06-SEP-2000: 2000US-0230437.
PR 06-SEP-2000: 2000US-0230438.
PR 08-SEP-2000: 2000US-0231242.
PR 08-SEP-2000: 2000US-0231243.
PR 08-SEP-2000: 2000US-0231244.
PR 08-SEP-2000: 2000US-0231413.
PR 08-SEP-2000: 2000US-0231414.
PR 08-SEP-2000: 2000US-0232080.
PR 08-SEP-2000: 2000US-0232081.
PR 12-SEP-2000: 2000US-0231968.
PR 14-SEP-2000: 2000US-0232397.
PR 14-SEP-2000: 2000US-0232398.
PR 14-SEP-2000: 2000US-0232399.
PR 14-SEP-2000: 2000US-0232400.
PR 14-SEP-2000: 2000US-0232401.
PR 14-SEP-2000: 2000US-0233063.
PR 14-SEP-2000: 2000US-0233064.
PR 14-SEP-2000: 2000US-0233065.
PR 21-SEP-2000: 2000US-0234223.
PR 21-SEP-2000: 2000US-0234274.
PR 25-SEP-2000: 2000US-0234597.
PR 25-SEP-2000: 2000US-0234598.
PR 26-SEP-2000: 2000US-0235484.
PR 27-SEP-2000: 2000US-0235834.
PR 27-SEP-2000: 2000US-0235835.
PR 29-SEP-2000: 2000US-0236327.
PR 29-SEP-2000: 2000US-0236367.
PR 29-SEP-2000: 2000US-0236368.
PR 29-SEP-2000: 2000US-0236369.
PR 29-SEP-2000: 2000US-0236370.
PR 02-OCT-2000: 2000US-0236802.
PR 02-OCT-2000: 2000US-0237037.
PR 02-OCT-2000: 2000US-0237038.
PR 02-OCT-2000: 2000US-0237039.
PR 02-OCT-2000: 2000US-0237040.
PR 13-OCT-2000: 2000US-0239935.
PR 13-OCT-2000: 2000US-0239937.
PR 20-OCT-2000: 2000US-0240960.
PR 20-OCT-2000: 2000US-0241221.
PR 20-OCT-2000: 2000US-0241785.
PR 20-OCT-2000: 2000US-0241786.
PR 20-OCT-2000: 2000US-0241787.
PR 20-OCT-2000: 2000US-0241808.
PR 20-OCT-2000: 2000US-0241809.
PR 20-OCT-2000: 2000US-0241826.
PR 01-NOV-2000: 2000US-0244617.
PR 08-NOV-2000: 2000US-0246474.
PR 08-NOV-2000: 2000US-0246475.
PR 08-NOV-2000: 2000US-0246476.
PR 08-NOV-2000: 2000US-0246477.
PR 08-NOV-2000: 2000US-0246478.
PR 08-NOV-2000: 2000US-0246523.
PR 08-NOV-2000: 2000US-0246524.
PR 08-NOV-2000: 2000US-0246525.
PR 08-NOV-2000: 2000US-0246526.
PR 08-NOV-2000: 2000US-0246527.
PR 08-NOV-2000: 2000US-0246528.
PR 08-NOV-2000: 2000US-0246532.
PR 08-NOV-2000: 2000US-0246539.
PR 08-NOV-2000: 2000US-0246610.
PR 08-NOV-2000: 2000US-0246611.
PR 08-NOV-2000: 2000US-0246613.
PR 17-NOV-2000: 2000US-0249207.
PR 17-NOV-2000: 2000US-0249208.
PR 17-NOV-2000: 2000US-0249209.
PR 17-NOV-2000: 2000US-0249210.
PR 17-NOV-2000: 2000US-0249211.
PR 17-NOV-2000: 2000US-0249212.
PR 17-NOV-2000: 2000US-0249213.

PR 17-NOV-2000: 2000US-0249214.
PR 17-NOV-2000: 2000US-0249215.
PR 17-NOV-2000: 2000US-0249216.
PR 17-NOV-2000: 2000US-0249217.
PR 17-NOV-2000: 2000US-0249218.
PR 17-NOV-2000: 2000US-0249244.
PR 17-NOV-2000: 2000US-0249245.
PR 17-NOV-2000: 2000US-0249264.
PR 17-NOV-2000: 2000US-0249265.
PR 17-NOV-2000: 2000US-0249297.
PR 17-NOV-2000: 2000US-0249299.
PR 17-NOV-2000: 2000US-0249300.
PR 01-DEC-2000: 2000US-0250160.
PR 01-DEC-2000: 2000US-0250391.
PR 05-DEC-2000: 2000US-0251030.
PR 05-DEC-2000: 2000US-0251988.
PR 05-DEC-2000: 2000US-0256179.
PR 06-DEC-2000: 2000US-0251479.
PR 08-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251858.
PR 08-DEC-2000: 2000US-0251869.
PR 08-DEC-2000: 2000US-0251989.
PR 08-DEC-2000: 2000US-0251990.
PR 11-DEC-2000: 2000US-0254097.
PR 05-JAN-2001: 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-451930/48.
DR N-PSDB: AAS35704.
XX
PT New cardiovascular system related polynucleotides and polypeptides,
PT useful for diagnosing, treating and/or preventing disorders of the
PT cardiovascular system -
XX
XX
PS Claim 11: SEQ ID NO 1204; 674bp; English.
XX
CC Sequences AMU21852-AMU22466 represent the cardiovascular system antigen
CC polypeptides of the invention. Cardiovascular system antigens and their
CC associated polynucleotides are useful in the diagnosis, treatment and
CC prevention of various types of disorders in e.g. humans, mice, rabbits,
CC goats, horses, cats, dogs, chickens or sheep. A pathological condition
CC can be determined by detecting the presence or absence of a mutation in a
CC cardiovascular system antigen polynucleotide. The treatable disorders
CC include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma and pleurisy. The polypeptides can
CC also be used to aid wound healing, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, to regenerate tissues and in
CC chemotaxis.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pcl_sequences](http://wipo.int/pub/published_pcl_sequences).

Query Match 0.8%; Score 7; DB 22; Length 51;

Best Local Similarity 100.0%; Pred. No. 89;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 LRGATLV 240

Db 25 LRGATLV 31

RESULT 31

AAM87002

ID AAM87002 standard; Protein: 62 AA.
XX AAM87002;
AC
DT 07-NOV-2001 (first entry)
XX
DE Human Immune/haematopoietic antigen SEQ ID NO:14595.
XX
KW Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytosolic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN W0200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01554.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0214886.
PR 28-JUN-2000; 2000US-0215135.
PR 30-JUN-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216688.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0220965.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225271.
PR 14-AUG-2000; 2000US-0225275.
PR 14-AUG-2000; 2000US-0225279.
PR 18-AUG-2000; 2000US-0225758.
PR 22-AUG-2000; 2000US-0225759.
PR 22-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0233297.
PR 14-SEP-2000; 2000US-0233298.

PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.

PR	05-DEC-2000;	2000US-0256719.		
PR	06-DEC-2000;	2000US-0251479.		
PR	08-DEC-2000;	2000US-0251856.		
PR	08-DEC-2000;	2000US-0251856.		
PR	08-DEC-2000;	2000US-0251868.		
PR	08-DEC-2000;	2000US-0251869.		
PR	08-DEC-2000;	2000US-0251989.		
PR	08-DEC-2000;	2000US-0251990.		
PR	11-DEC-2000;	2000US-0254097.		
PR	05-JAN-2001;	2001US-0259678.		
PA	(HUMA-) HUMAN GENOME SCI INC.			
PI	Rosen CA,	Barash SC, Ruben SM;		
DR	WPI: 2001-483426/52.			
XX	N-PSDB; AAK59783.			
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,			
PT	useful for preventing, diagnosing and/or treating cancers and			
PT	metastasis -			
XX	Claim 11; SEQ ID NO 14595; 3071bp + Sequence Listing; English.			
XX	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)			
CC	amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic			
CC	activity, and can be used in gene therapy and vaccine production. (I)			
CC	proteins and polynucleotides may be used in the prevention, diagnosis and			
CC	treatment of diseases associated with inappropriate (I) expression. For			
CC	example, they may be used to treat disorders associated with decreased			
CC	expression by rectifying mutations or deletions in a patient's genome			
CC	that affect the activity of (I) by expressing inactive proteins or to			
CC	supplement the patient's own production of (I). Additionally, (I)			
CC	polynucleotides may be used to produce the secreted (I), by inserting the			
CC	nucleic acids into a host cell and culturing the cell to express the			
CC	protein. (I) proteins and polynucleotides may be used to prevent,			
CC	diagnose and treat immune/hematopoietic-related diseases, especially			
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703			
CC	to AAK67694 represent human immune/hematopoietic antigen genomic			
CC	sequences from the present invention. AAK54942 to AAK54950 and AAK62169			
XX	represent sequences used in the exemplification of the present invention.			
SO	Sequence 62 AA:			
Query Match 0.8%; Score 7; DB 22; Length 62;				
Best Local Similarity 100.0%; Pred. No. 1.1e+02;				
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0				
OY	407 VRALSND 413			
DB	39 VRALSND 45			
RESULT 32				
AY11419				
ID	AA11419 standard; Protein: 64 AA.			
XX	AA11419;			
XX	21-JUN-1999 (first entry)			
DT	Human 5' EST secreted protein SEQ ID NO 241.			
DE				
XX	Human; secreted protein; EST; expressed sequence tag; diagnosis;			
KM	forensic; gene therapy; chromosome mapping; signal peptide;			
KM	upstream regulatory sequence; cytokine activity; cell proliferation;			
KM	differentiation; haematopoiesis regulation; tissue growth regulation;			
KM	reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;			
XX	thrombolytic; anti-inflammatory; tumour inhibition.			
OS	Homo sapiens.			
XX	W0906551-A2.			

```

PD      11-FEB-1999.
PF      31-JUL-1998;    98WO-IB01235.
XX
PR      01-AUG-1997;    9TUS-0905133.
XX
PA      (GEST ) GENSET.
XX
PI      Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR      MPI: 1999-153781/13.
XX      N-PSTDB; AAX39485.
XX
PT      New nucleic acids encoding human secreted - proteins obtained from
XX      CDNA libraries prepared from substantia nigra, cerebellum, surrenals
XX      and fetal brain tissue
XX
PS      Claim 34; Page 364; 434pp; English.
XX
CC      AAX39440 to AAX39597 represent 5' expressed sequence tags (ESTs) for
XX      human secreted proteins, and encode the proteins given in AAU1374 to
XX      CC AAU1531, respectively. The proteins given represent the signal peptide
XX      and an N-terminal fragment of a secreted protein. The nucleic acid
XX      sequences can be used for producing secreted human gene products. They
XX      can also be used to develop products for diagnosis and therapy. The
XX      CC proteins obtained may have cytokine activity, cell
XX      proliferation/differentiation activity, haematopoiesis regulating
XX      CC activity, tissue growth regulating activity, reproductive hormone
XX      regulating activity, chemotactic/chemokinetic activity, haemostatic and
XX      thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX      CC activity, tumour inhibition activity or other activities. The products
XX      can be used in forensic, gene therapy and chromosome mapping procedures.
XX      The sequences can also be used for obtaining corresponding promoter
XX      CC sequences. The nucleic acids encoding the signal peptide can be used for
XX      directing extracellular secretion of a polypeptide or the insertion of a
XX      polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ      Sequence      64 AA:

Query Match          0.8%; Score 7; DB 20; Length 64;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      573 LVFVVAS 579
        |||||
Db       17 LVFVVAS 23

RESULT 33
AAAG77071
ID      AAG77071 standard; Protein: 64 AA.
XX
AC      AAG77071;
XX
DT      03-SEP-2001 (first entry)
XX
DE      Human colon cancer antigen protein SEQ ID NO:7835.
XX
KW      Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX      colorectal carcinoma.
XX
OS      Homo sapiens.
XX
PN      WO200122920-A2.
XX
PD      05-APR-2001.
XX
PF      28-SEP-2000; 2000WO-US26524.
XX
PR      29-SEP-1999;    99US-0157137.
XX      03-NOV-1999;    99US-0163280.
XX
(PUMA-) HUMAN GENOME SCI INC.
```



```

PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057669.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
PI Feng P, Ferlie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX
XX WPI: 1998-506364/43.
DR N-PSDB: AAV59785.
XX
XX New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
XX Claim 1: Page 694; 721pp; English.
XX
XX This sequence represents a secreted human protein encoded by the nucleic
CC acid molecule designated Gene 146 from the human cDNA clone HSNK17
CC (deposited as clone ATCC 97903 and ATCC 209049).
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. AAV59502) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: AAV59511-V59812; amino acid sequences AAV74731-W5026)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 186
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAV59511 for described uses).
XX
XX Sequence 69 AA:
SQ

```

```

Query Match 0.8%; Score 7; DB 19; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 388 ELLEQLL 394
    |||||
DB 51 ELLEQLL 57

```

```

RESULT 35
AAAG01031
ID AAAG01031 standard; Protein; 69 AA.
XX
XX AAAG01031;
AC
XX
XX 06-OCT-2000 (first entry)
DT
XX
XX Human secreted protein, SEQ ID NO: 5112.
DE
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
XX
XX

```

```

XX
XX Homo sapiens.
OS
XX
XX EP1033401-A2.
XX
XX
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-0200610.
PE
XX
XX 26-FEB-1999; 99US-0122487.
PR
XX
XX (GSEST ) GENSET.
PA
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX
XX WPI: 2000-500381/45.
DR N-PSDB: AAC01037.
XX
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PT
XX
XX Claim 13; SEQ ID 5112; 71pp + CD-ROM; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC regulatory mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
XX Sequence 69 AA:
SQ

```

```

Query Match 0.8%; Score 7; DB 21; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 482 FSASNLI 488
    |||||
DB 59 FSASNLI 65

```

```

RESULT 36
AAAG01032
ID AAAG01032 standard; Protein; 69 AA.
XX
XX AAAG01032;
AC
XX
XX 06-OCT-2000 (first entry)
DT
XX
XX Human secreted protein, SEQ ID NO: 5113.
DE
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
XX Homo sapiens.
OS
XX
XX EP1033401-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-0200610.
PE
XX
XX 26-FEB-1999; 99US-0122487.
PR
XX
XX (GSEST ) GENSET.
PA

```

```
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI: 2000-500381/45.
DR N-PSDB: AAC01038.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13: SEQ ID 5113; 71bp + CD-ROM; English.
XX
CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or poly(A) RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
SO Sequence 69 AA:
Query Match 0.8%; Score 7; DB 21; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 482 FSASNLI 488
Db 59 FSASNLI 65
|||||
RESULT 37
AAW74874
ID AAW74874 standard; Protein: 70 AA.
XX
AC AAW74874;
XX
DT 25-JAN-1999 (first entry)
XX
DE Human secreted protein encoded by gene 146 clone HSNK17.
XX
DE Human; secreted protein; testis; tumour; foetal brain tissue;
XX fusion protein; cancer; central nervous system; seizure;
XX diagnosis; neurodegenerative disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 70 /label= unknown
XX
PN WC9839448-A2.
XX
PD 11-SEP-1998.
XX
PF 06-MAR-1998; 98MO-US04493.
XX
XX 02-OCT-1997; 97US-0061060.
XX 07-MAR-1997; 97US-0038621.
XX 07-MAR-1997; 97US-0040161.
XX 07-MAR-1997; 97US-0040162.
XX 07-MAR-1997; 97US-0040163.
XX 07-MAR-1997; 97US-0040334.
XX 07-MAR-1997; 97US-0040336.
XX 07-MAR-1997; 97US-0040626.
XX 11-APR-1997; 97US-0043311.
```

```
PR 11-APR-1997; 97US-0043312.
PR 11-APR-1997; 97US-0043313.
PR 11-APR-1997; 97US-0043314.
PR 11-APR-1997; 97US-0043568.
PR 11-APR-1997; 97US-0043569.
PR 11-APR-1997; 97US-0043576.
PR 11-APR-1997; 97US-0043578.
PR 11-APR-1997; 97US-0043580.
PR 11-APR-1997; 97US-0043669.
PR 11-APR-1997; 97US-0043670.
PR 11-APR-1997; 97US-0043671.
PR 11-APR-1997; 97US-0043672.
PR 11-APR-1997; 97US-0043674.
PR 11-APR-1997; 97US-0043676.
PR 11-APR-1997; 97US-0043678.
PR 11-APR-1997; 97US-0043680.
PR 11-APR-1997; 97US-0043682.
PR 11-APR-1997; 97US-0043684.
PR 11-APR-1997; 97US-0043686.
PR 11-APR-1997; 97US-0043688.
PR 11-APR-1997; 97US-0043690.
PR 11-APR-1997; 97US-0043692.
PR 11-APR-1997; 97US-0043694.
PR 11-APR-1997; 97US-0043696.
PR 11-APR-1997; 97US-0043698.
PR 11-APR-1997; 97US-0043700.
PR 11-APR-1997; 97US-0043702.
PR 11-APR-1997; 97US-0043704.
PR 11-APR-1997; 97US-0043706.
PR 11-APR-1997; 97US-0043708.
PR 11-APR-1997; 97US-0043710.
PR 11-APR-1997; 97US-0043712.
PR 11-APR-1997; 97US-0043714.
PR 11-APR-1997; 97US-0043716.
PR 11-APR-1997; 97US-0043718.
PR 11-APR-1997; 97US-0043720.
PR 11-APR-1997; 97US-0043722.
PR 11-APR-1997; 97US-0043724.
PR 11-APR-1997; 97US-0043726.
PR 11-APR-1997; 97US-0043728.
PR 11-APR-1997; 97US-0043730.
PR 11-APR-1997; 97US-0043732.
PR 11-APR-1997; 97US-0043734.
PR 11-APR-1997; 97US-0043736.
PR 11-APR-1997; 97US-0043738.
PR 11-APR-1997; 97US-0043740.
PR 11-APR-1997; 97US-0043742.
PR 11-APR-1997; 97US-0043744.
PR 11-APR-1997; 97US-0043746.
PR 11-APR-1997; 97US-0043748.
PR 11-APR-1997; 97US-0043750.
PR 11-APR-1997; 97US-0043752.
PR 11-APR-1997; 97US-0043754.
PR 11-APR-1997; 97US-0043756.
PR 11-APR-1997; 97US-0043758.
PR 11-APR-1997; 97US-0043760.
PR 11-APR-1997; 97US-0043762.
PR 11-APR-1997; 97US-0043764.
PR 11-APR-1997; 97US-0043766.
PR 11-APR-1997; 97US-0043768.
PR 11-APR-1997; 97US-0043770.
PR 11-APR-1997; 97US-0043772.
PR 11-APR-1997; 97US-0043774.
PR 11-APR-1997; 97US-0043776.
PR 11-APR-1997; 97US-0043778.
PR 11-APR-1997; 97US-0043780.
PR 11-APR-1997; 97US-0043782.
PR 11-APR-1997; 97US-0043784.
PR 11-APR-1997; 97US-0043786.
PR 11-APR-1997; 97US-0043788.
PR 11-APR-1997; 97US-0043790.
PR 11-APR-1997; 97US-0043792.
PR 11-APR-1997; 97US-0043794.
PR 11-APR-1997; 97US-0043796.
PR 11-APR-1997; 97US-0043798.
PR 11-APR-1997; 97US-0043800.
PR 11-APR-1997; 97US-0043802.
PR 11-APR-1997; 97US-0043804.
PR 11-APR-1997; 97US-0043806.
PR 11-APR-1997; 97US-0043808.
PR 11-APR-1997; 97US-0043810.
PR 11-APR-1997; 97US-0043812.
PR 11-APR-1997; 97US-0043814.
PR 11-APR-1997; 97US-0043816.
PR 11-APR-1997; 97US-0043818.
PR 11-APR-1997; 97US-0043820.
PR 11-APR-1997; 97US-0043822.
PR 11-APR-1997; 97US-0043824.
PR 11-APR-1997; 97US-0043826.
PR 11-APR-1997; 97US-0043828.
PR 11-APR-1997; 97US-0043830.
PR 11-APR-1997; 97US-0043832.
PR 11-APR-1997; 97US-0043834.
PR 11-APR-1997; 97US-0043836.
PR 11-APR-1997; 97US-0043838.
PR 11-APR-1997; 97US-0043840.
PR 11-APR-1997; 97US-0043842.
PR 11-APR-1997; 97US-0043844.
PR 11-APR-1997; 97US-0043846.
PR 11-APR-1997; 97US-0043848.
PR 11-APR-1997; 97US-0043850.
PR 11-APR-1997; 97US-0043852.
PR 11-APR-1997; 97US-0043854.
PR 11-APR-1997; 97US-0043856.
PR 11-APR-1997; 97US-0043858.
PR 11-APR-1997; 97US-0043860.
PR 11-APR-1997; 97US-0043862.
PR 11-APR-1997; 97US-0043864.
PR 11-APR-1997; 97US-0043866.
PR 11-APR-1997; 97US-0043868.
PR 11-APR-1997; 97US-0043870.
PR 11-APR-1997; 97US-0043872.
PR 11-APR-1997; 97US-0043874.
PR 11-APR-1997; 97US-0043876.
PR 11-APR-1997; 97US-0043878.
PR 11-APR-1997; 97US-0043880.
PR 11-APR-1997; 97US-0043882.
PR 11-APR-1997; 97US-0043884.
PR 11-APR-1997; 97US-0043886.
PR 11-APR-1997; 97US-0043888.
PR 11-APR-1997; 97US-0043890.
PR 11-APR-1997; 97US-0043892.
PR 11-APR-1997; 97US-0043894.
PR 11-APR-1997; 97US-0043896.
PR 11-APR-1997; 97US-0043898.
PR 11-APR-1997; 97US-0043900.
PR 11-APR-1997; 97US-0043902.
PR 11-APR-1997; 97US-0043904.
PR 11-APR-1997; 97US-0043906.
PR 11-APR-1997; 97US-0043908.
PR 11-APR-1997; 97US-0043910.
PR 11-APR-1997; 97US-0043912.
PR 11-APR-1997; 97US-0043914.
PR 11-APR-1997; 97US-0043916.
PR 11-APR-1997; 97US-0043918.
PR 11-APR-1997; 97US-0043920.
PR 11-APR-1997; 97US-0043922.
PR 11-APR-1997; 97US-0043924.
PR 11-APR-1997; 97US-0043926.
PR 11-APR-1997; 97US-0043928.
PR 11-APR-1997; 97US-0043930.
PR 11-APR-1997; 97US-0043932.
PR 11-APR-1997; 97US-0043934.
PR 11-APR-1997; 97US-0043936.
PR 11-APR-1997; 97US-0043938.
PR 11-APR-1997; 97US-0043940.
PR 11-APR-1997; 97US-0043942.
PR 11-APR-1997; 97US-0043944.
PR 11-APR-1997; 97US-0043946.
PR 11-APR-1997; 97US-0043948.
PR 11-APR-1997; 97US-0043950.
PR 11-APR-1997; 97US-0043952.
PR 11-APR-1997; 97US-0043954.
PR 11-APR-1997; 97US-0043956.
PR 11-APR-1997; 97US-0043958.
PR 11-APR-1997; 97US-0043960.
PR 11-APR-1997; 97US-0043962.
PR 11-APR-1997; 97US-0043964.
PR 11-APR-1997; 97US-0043966.
PR 11-APR-1997; 97US-0043968.
PR 11-APR-1997; 97US-0043970.
PR 11-APR-1997; 97US-0043972.
PR 11-APR-1997; 97US-0043974.
PR 11-APR-1997; 97US-0043976.
PR 11-APR-1997; 97US-0043978.
PR 11-APR-1997; 97US-0043980.
PR 11-APR-1997; 97US-0043982.
PR 11-APR-1997; 97US-0043984.
PR 11-APR-1997; 97US-0043986.
PR 11-APR-1997; 97US-0043988.
PR 11-APR-1997; 97US-0043990.
PR 11-APR-1997; 97US-0043992.
PR 11-APR-1997; 97US-0043994.
PR 11-APR-1997; 97US-0043996.
PR 11-APR-1997; 97US-0043998.
PR 11-APR-1997; 97US-0044000.
PR 11-APR-1997; 97US-0044002.
PR 11-APR-1997; 97US-0044004.
PR 11-APR-1997; 97US-0044006.
PR 11-APR-1997; 97US-0044008.
PR 11-APR-1997; 97US-0044010.
PR 11-APR-1997; 97US-0044012.
PR 11-APR-1997; 97US-0044014.
PR 11-APR-1997; 97US-0044016.
PR 11-APR-1997; 97US-0044018.
PR 11-APR-1997; 97US-0044020.
PR 11-APR-1997; 97US-0044022.
PR 11-APR-1997; 97US-0044024.
PR 11-APR-1997; 97US-0044026.
PR 11-APR-1997; 97US-0044028.
PR 11-APR-1997; 97US-0044030.
PR 11-APR-1997; 97US-0044032.
PR 11-APR-1997; 97US-0044034.
PR 11-APR-1997; 97US-0044036.
PR 11-APR-1997; 97US-0044038.
PR 11-APR-1997; 97US-0044040.
PR 11-APR-1997; 97US-0044042.
PR 11-APR-1997; 97US-0044044.
PR 11-APR-1997; 97US-0044046.
PR 11-APR-1997; 97US-0044048.
PR 11-APR-1997; 97US-0044050.
PR 11-APR-1997; 97US-0044052.
PR 11-APR-1997; 97US-0044054.
PR 11-APR-1997; 97US-0044056.
PR 11-APR-1997; 97US-0044058.
PR 11-APR-1997; 97US-0044060.
PR 11-APR-1997; 97US-0044062.
PR 11-APR-1997; 97US-0044064.
PR 11-APR-1997; 97US-0044066.
PR 11-APR-1997; 97US-0044068.
PR 11-APR-1997; 97US-0044070.
PR 11-APR-1997; 97US-0044072.
PR 11-APR-1997; 97US-0044074.
PR 11-APR-1997; 97US-0044076.
PR 11-APR-1997; 97US-0044078.
PR 11-APR-1997; 97US-0044080.
PR 11-APR-1997; 97US-0044082.
PR 11-APR-1997; 97US-0044084.
PR 11-APR-1997; 97US-0044086.
PR 11-APR-1997; 97US-0044088.
PR 11-APR-1997; 97US-0044090.
PR 11-APR-1997; 97US-0044092.
PR 11-APR-1997; 97US-0044094.
PR 11-APR-1997; 97US-0044096.
PR 11-APR-1997; 97US-0044098.
PR 11-APR-1997; 97US-0044100.
PR 11-APR-1997; 97US-0044102.
PR 11-APR-1997; 97US-0044104.
PR 11-APR-1997; 97US-0044106.
PR 11-APR-1997; 97US-0044108.
PR 11-APR-1997; 97US-0044110.
PR 11-APR-1997; 97US-0044112.
PR 11-APR-1997; 97US-0044114.
PR 11-APR-1997; 97US-0044116.
PR 11-APR-1997; 97US-0044118.
PR 11-APR-1997; 97US-0044120.
PR 11-APR-1997; 97US-0044122.
PR 11-APR-1997; 97US-0044124.
PR 11-APR-1997; 97US-0044126.
PR 11-APR-1997; 97US-0044128.
PR 11-APR-1997; 97US-0044130.
PR 11-APR-1997; 97US-0044132.
PR 11-APR-1997; 97US-0044134.
PR 11-APR-1997; 97US-0044136.
PR 11-APR-1997; 97US-0044138.
PR 11-APR-1997; 97US-0044140.
PR 11-APR-1997; 97US-0044142.
PR 11-APR-1997; 97US-0044144.
PR 11-APR-1997; 97US-0044146.
PR 11-APR-1997; 97US-0044148.
PR 11-APR-1997; 97US-0044150.
PR 11-APR-1997; 97US-0044152.
PR 11-APR-1997; 97US-0044154.
PR 11-APR-1997; 97US-0044156.
PR 11-APR-1997; 97US-0044158.
PR 11-APR-1997; 97US-0044160.
PR 11-APR-1997; 97US-0044162.
PR 11-APR-1997; 97US-0044164.
PR 11-APR-1997; 97US-0044166.
PR 11-APR-1997; 97US-0044168.
PR 11-APR-1997; 97US-0044170.
PR 11-APR-1997; 97US-0044172.
PR 11-APR-1997; 97US-0044174.
PR 11-APR-1997; 97US-0044176.
PR 11-APR-1997; 97US-0044178.
PR 11-APR-1997; 97US-0044180.
PR 11-APR-1997; 97US-0044182.
PR 11-APR-1997; 97US-0044184.
PR 11-APR-1997; 97US-0044186.
PR 11-APR-1997; 97US-0044188.
PR 11-APR-1997; 97US-0044190.
PR 11-APR-1997; 97US-0044192.
PR 11-APR-1997; 97US-0044194.
PR 11-APR-1997; 97US-0044196.
PR 11-APR-1997; 97US-0044198.
PR 11-APR-1997; 97US-0044200.
PR 11-APR-1997; 97US-0044202.
PR 11-APR-1997; 97US-0044204.
PR 11-APR-1997; 97US-0044206.
PR 11-APR-1997; 97US-0044208.
PR 11-APR-1997; 97US-0044210.
PR 11-APR-1997; 97US-0044212.
PR 11-APR-1997; 97US-0044214.
PR 11-APR-1997; 97US-0044216.
PR 11-APR-1997; 97US-0044218.
PR 11-APR-1997; 97US-0044220.
PR 11-APR-1997; 97US-0044222.
PR 11-APR-1997; 97US-0044224.
PR 11-APR-1997; 97US-0044226.
PR 11-APR-1997; 97US-0044228.
PR 11-APR-1997; 97US-0044230.
PR 11-APR-1997; 97US-0044232.
PR 11-APR-1997; 97US-0044234.
PR 11-APR-1997; 97US-0044236.
PR 11-APR-1997; 97US-0044238.
PR 11-APR-1997; 97US-0044240.
PR 11-APR-1997; 97US-0044242.
PR 11-APR-1997; 97US-0044244.
PR 11-APR-1997; 97US-0044246.
PR 11-APR-1997; 97US-0044248.
PR 11-APR-1997; 97US-0044250.
PR 11-APR-1997; 97US-0044252.
PR 11-APR-1997; 97US-0044254.
PR 11-APR-1997; 97US-0044256.
PR 11-APR-1997; 97US-0044258.
PR 11-APR-1997; 97US-0044260.
PR 11-APR-1997; 97US-0044262.
PR 11-APR-1997; 97US-0044264.
PR 11-APR-1997; 97US-0044266.
PR 11-APR-1997; 97US-0044268.
PR 11-APR-1997; 97US-0044270.
PR 11-APR-1997; 97US-0044272.
PR 11-APR-1997; 97US-0044274.
PR 11-APR-1997; 97US-0044276.
PR 11-APR-1997; 97US-0044278.
PR 11-APR-1997; 97US-0044280.
PR 11-APR-1997; 97US-0044282.
PR 11-APR-1997; 97US-0044284.
PR 11-APR-1997; 97US-0044286.
PR 11-APR-1997; 97US-0044288.
PR 11-APR-1997; 97US-0044290.
PR 11-APR-1997; 97US-0044292.
PR 11-APR-1997; 97US-0044294.
PR 11-APR-1997; 97US-0044296.
PR 11-APR-1997; 97US-0044298.
PR 11-APR-1997; 97US-0044300.
PR 11-APR-1997; 97US-0044302.
PR 11-APR-1997; 97US-0044304.
PR 11-APR-1997; 97US-0044306.
PR 11-APR-1997; 97US-0044308.
PR 11-APR-1997; 97US-0044310.
PR 11-APR-1997; 97US-0044312.
PR 11-APR-1997; 97US-0044314.
PR 11-APR-1997; 97US-0044316.
PR 11-APR-1997; 97US-0044318.
PR 11-APR-1997; 97US-0044320.
PR 11-APR-1997; 97US-0044322.
PR 11-APR-1997; 97US-0044324.
PR 11-APR-1997; 97US-0044326.
PR 11-APR-1997; 97US-0044328.
PR 11-APR-1997; 97US-0044330.
PR 11-APR-1997; 97US-0044332.
PR 11-APR-1997; 97US-0044334.
PR 11-APR-1997; 97US-0044336.
PR 11-APR-1997; 97US-0044338.
PR 11-APR-1997; 97US-0044340.
PR 11-APR-1997; 97US-0044342.
PR 11-APR-1997; 97US-0044344.
PR 11-APR-1997; 97US-0044346.
PR 11-APR-1997; 97US-0044348.
PR 11-APR-1997; 97US-0044350.
PR 11-APR-1997; 97US-0044352.
PR 11-APR-1997; 97US-0044354.
PR 11-APR-1997; 97US-0044356.
PR 11-APR-1997; 97US-0044358.
PR 11-APR-1997; 97US-0044360.
PR 11-APR-1997; 97US-0044362.
PR 11-APR-1997; 97US-0044364.
PR 11-APR-1997; 97US-0044366.
PR 11-APR-1997; 97US-0044368.
PR 11-APR-1997; 97US-0044370.
PR 11-APR-1997; 97US-0044372.
PR 11-APR-1997; 97US-0044374.
PR 11-APR-1997; 97US-0044376.
PR 11-APR-1997; 97US-0044378.
PR 11-APR-1997; 97US-0044380.
PR 11-APR-1997; 97US-0044382.
PR 11-APR-1997; 97US-0044384.
PR 11-APR-1997; 97US-0044386.
PR 11-APR-1997; 97US-0044388.
PR 11-APR-1997; 97US-0044390.
PR 11-APR-1997; 97US-0044392.
PR 11-APR-1997; 97US-0044394.
PR 11-APR-1997; 97US-0044396.
PR 11-APR-1997; 97US-0044398.
PR 11-APR-1997; 97US-0044400.
PR 11-APR-1997; 97US-0044402.
PR 11-APR-1997; 97US-0044404.
PR 11-APR-1997; 97US-0044406.
PR 11-APR-1997; 97US-0044408.
PR 11-APR-1997; 97US-0044410.
PR 11-APR-1997; 97US-0044412.
PR 11-APR-1997; 97US-0044414.
PR 11-APR-1997; 97US-0044416.
PR 11-APR-1997; 97US-0044418.
PR 11-APR-1997; 97US-0044420.
PR 11-APR-1997; 97US-0044422.
PR 11-APR-1997; 97US-0044424.
PR 11-APR-1997; 97US-0044426.
PR 11-APR-1997; 97US-0044428.
PR 11-APR-1997; 97US-0044430.
PR 11-APR-1997; 97US-0044432.
PR 11-APR-1997; 97US-0044434.
PR 11-APR-1997; 97US-0044436.
PR 11-APR-1997; 97US-0044438.
PR 11-APR-1997; 97US-0044440.
PR 11-APR-1997; 97US-0044442.
PR 11-APR-1997; 97US-0044444.
PR 11-APR-1997; 97US-0044446.
PR 11-APR-1997; 97US-0044448.
PR 11-APR-1997; 97US-0044450.
PR 11-APR-1997; 97US-0044452.
PR 11-APR-1997; 97US-0044454.
PR 11-APR-1997; 97US-0044456.
PR 11-APR-1997; 97US-0044458.
PR 11-APR-1997; 97US-0044460.
PR 11-APR-1997; 97US-0044462.
PR 11-APR-1997; 97US-0044464.
PR 11-APR-1997; 97US-0044466.
PR 11-APR-1997; 97US-0044468.
PR 11-APR-1997; 97US-0044470.
PR 11-APR-1997; 97US-0044472.
PR 11-APR-1997; 97US-0044474.
PR 11-APR-1997; 97US-0044476.
PR 11-APR-1997; 97US-0044478.
PR 11-APR-1997; 97US-0044480.
PR 11-APR-1997; 97US-0044482
```

PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057669.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
PI Feng P, Ferrie AM, Fischer CL, Florence KM, Greene JM, Hu JS;
PI Kyaw H, Lafleur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX
XX WPI: 1998-506364/43.
DR N-PSDB: AAV59656.
XX
XX
XX New isolated human genes and the secreted polypeptide(s) they encode
XX - useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders
XX
XX
XX Claim 1; Page 628; 721pp; English.
XX
XX
XX This sequence represents a secreted human protein encoded by the nucleic
XX acid molecule designated Gene 146 from the human cDNA clone HSNM17
XX (deposited as clone ATCC 97903 and ATCC 209049).
XX The gene can be used to generate fusion proteins by linking to the gene
XX to a human immunoglobulin Fc portion (e.g. AAV59502) for increasing the
XX stability of the fused protein as compared to the human protein only.
XX The invention relates to 186 novel genes and their fragments (nucleic
XX acid sequences: AAV59511-V59812; amino acid sequences AAW4731-W5026)
XX which are useful for preventing, treating or ameliorating medical
XX conditions e.g. by protein or gene therapy. Also, pathological
XX conditions can be diagnosed by determining the amount of the new
XX polypeptides in a sample or by determining the presence of mutations in
XX the new polynucleotides. Specific uses are described for each of the 186
XX polynucleotides, based on which tissues they are most highly expressed in
XX (see AAV59511 for described uses).
XX
XX
XX Sequence 70 AA:
SQ
Query Match 0.8%; Score 7; DB 19; Length 70;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 388 ELLLEQL 394
DB 51 ELLLEQL 57
RESULT 38
ABP05664
ID ABP05664 standard; Protein; 79 AA.
XX
XX ABP05664;
XX
XX 24-JUN-2002 (first entry)
XX
XX Human ORFX protein sequence SEQ ID NO:11310.
XX
XX Human: open reading frame; ORFX: gene therapy; cancer; cirrhosis;
XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
XX degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW

KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
XX Homo sapiens.
XX
XX WO200192523-A2.
XX
XX 06-DEC-2001.
XX
XX
XX 29-MAY-2001; 2001WO-US10836.
XX
XX 30-MAY-2000; 2000US-206132P.
XX
XX 29-AUG-2000; 2000US-228716P.
XX
XX
XX (CURA-) CURAGEN CORP.
XX
XX
XX Shinkets RA, Leach MD;
XX
XX WPI: 2002-106308/14.
XX
XX N-PSDB: ABN21416.
XX
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and autoimmune disorders
XX
XX
XX Disclosure; SEQ ID 11310; 1037pp; English.
XX
XX
XX The present invention describes substantially purified human proteins
XX (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
XX in the specification)). ABN15762 to ABN27252 encode the human ORFX
XX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
XX treating or preventing a pathology associated with an ORFX-associated
XX disorder in humans, and in the manufacture of a medicament for treating a
XX syndrome associated with ORFX-associated disorder. ORFX polynucleotide
XX sequences can be used in gene therapy. ORFX sequences can be used in the
XX treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
XX psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,
XX osteoarthritis, neurodegenerative disorders, disorders related to organ
XX transplantation, cardiovascular diseases, diabetes mellitus, systemic
XX lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
XX storage disease, various immune deficiencies and disorders, infectious
XX diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
XX arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
XX disease and autoimmune inflammatory eye disease. ORFX proteins are also
XX useful for treating burns, incisions, ulcers, for treating osteoporosis,
XX bone degenerative disorders, or periodontal disease, and for gut
XX protection or regeneration and treatment of lung or liver fibrosis,
XX reperfusion injury in various tissues and conditions resulting from
XX systemic cytokine damage.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 79 AA:
SQ
Query Match 0.8%; Score 7; DB 23; Length 79;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 387 RELLEQL 393
DB 56 RELLEQL 62
RESULT 39
ABBI6701
ID ABBI6701 standard; Protein; 82 AA.
XX
XX ABBI6701;
XX
XX

DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polypeptide SEQ ID NO 5358.
XX
XX Human: nootropic; neuroprotective; cyrostatic; dermatological; virucide;
KW immunosuppressive; antineoplastic; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antischiz; antidiabetic; antileukemic; cancer;
KW antileukemic; hepatotropic; cerebroprotective; antileukemic;
KW antileukemic; antidiabetic; antileukemic; antileukemic;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine.
OS Homo sapiens.
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0186874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0209467.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0225759.
PR 22-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227709.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0242221.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000: 2000US-0251988.
 PR 05-DEC-2000: 2000US-0256719.
 PR 06-DEC-2000: 2000US-0251479.
 PR 08-DEC-2000: 2000US-0251856.
 PR 08-DEC-2000: 2000US-0251868.
 PR 08-DEC-2000: 2000US-0251869.
 PR 08-DEC-2000: 2000US-0251989.
 PR 08-DEC-2000: 2000US-0251990.
 PR 11-DEC-2000: 2000US-0254097.
 PR 05-JAN-2001: 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI: 2001-541565/60.
 DR N-PSDB: ABA13027.
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 PS Claim 11: SEQ ID NO 5358; 1701bp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
 CC (AB14678-AB18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 82 AA:
 Query Match 0.8%; Score 7; DB 22; Length 82;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 752 GSSGLVS 758
 DB 69 GSSGLVS 75
 RESULT 40
 AAG98900
 ID AAG98900 standard; Protein: 82 AA.
 AC AAG98900;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE E. coli growth and proliferation related protein sequence SEQ ID NO:370.
 XX
 KW Escherichia coli: growth; proliferation; microbial; antimicrobial;
 KW bacterial infection; microorganism.
 XX
 OS Escherichia coli.
 OS
 PN WO200134810-A2.
 XX
 PD 17-MAY-2001.
 XX

PF 09-NOV-2000: 2000WO-US30950.
 XX
 PR 09-NOV-1999: 99US-0164415.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Forsyth RA, Ohlsen K, Zyskind J;
 DR WPI: 2001-335933/35.
 DR N-PSDB: AAH84571.
 XX
 PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful
 PT for screening for homologous genes and for designing expression vectors
 PT -
 XX
 PS Claim 19: Page 442; 522zp; English.
 XX
 CC AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
 CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli
 CC growth and proliferation related proteins given in AAG99078 and AAG98830
 CC to AAG98999. (I) can be used as potential targets for the generation of
 CC new antimicrobial agents, and for identification of compounds which
 CC interact with the gene products of (I). In addition the expression of
 CC (I) and the purification of the proteins, the purified proteins can be
 CC used to generate reagents and screen small molecule libraries or other
 CC candidate compound libraries for compounds that can be further developed
 CC to yield novel antimicrobial compounds. In addition, nucleic acid probes
 CC complementary to (I) that are specific for particular species of
 CC microorganisms can be used to identify particular microorganism species
 CC in clinical specimens, therefore, providing a rapid and dependable
 CC method by which to identify the causative agents of a bacterial
 CC infection. Also, antibodies generated against proteins translated from
 CC mRNA transcribed from proliferation-required sequences can also be used
 CC to screen for specific microorganisms that produce such proteins in a
 CC species-specific manner. AAH84371 and AAH84670 represent sequencing
 CC primers used in the isolation of E. coli growth and proliferation
 CC related sequence, which are used in an example from the present
 CC invention.
 XX
 SQ Sequence 82 AA:
 Query Match 0.8%; Score 7; DB 22; Length 82;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 839 GQVLAVG 845
 DB 7 GQVLAVG 13

Search completed: April 28, 2003, 16:33:09
 Job time : 97 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:32:45 ; Search time 19 Seconds

(without alignments)
1423.140 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 919
Sequence: 1 MSKPYLFANRSEFPVALAA.....TGVKEGPNPIKHFITGTPF 919Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/prodata/1/1aa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/1/1aa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/1/1aa/6A_COMB.pep.*
- 4: /cgn2_6/prodata/1/1aa/6B_COMB.pep.*
- 5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep.*
- 6: /cgn2_6/prodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.0	851	1 US-08-369-796-2	Sequence 2, Appli
2	9	1.0	851	2 US-08-852-091-2	Sequence 2, Appli
3	9	1.0	851	2 US-08-820-754-2	Sequence 2, Appli
4	9	1.0	851	3 US-08-956-652-2	Sequence 2, Appli
5	9	1.0	851	3 US-08-956-869-2	Sequence 2, Appli
6	9	1.0	851	3 US-09-012-710-2	Sequence 2, Appli
7	9	1.0	851	3 US-08-948-547-2	Sequence 2, Appli
8	9	1.0	851	4 US-09-087-465-4	Sequence 2, Appli
9	9	1.0	851	4 US-09-364-970-2	Sequence 2, Appli
10	9	1.0	851	4 US-09-566-273-2	Sequence 2, Appli
11	9	1.0	851	4 US-08-956-653A-2	Sequence 2, Appli
12	9	1.0	851	5 PCT-US95-17025-2	Sequence 2, Appli
13	9	1.0	852	1 US-08-276-099A-13	Sequence 13, Appli
14	9	1.0	852	1 US-08-781-890-13	Sequence 13, Appli
15	8	0.9	582	4 US-09-242-632A-2	Sequence 2, Appli
16	7	0.8	29	1 US-07-960-510-4	Sequence 5, Appli
17	7	0.8	29	1 US-07-960-510-5	Sequence 5, Appli
18	7	0.8	29	1 US-07-960-510-6	Sequence 6, Appli
19	7	0.8	29	1 US-07-960-510-7	Sequence 7, Appli
20	7	0.8	34	6 5187089-8	Patent No. 5187089
21	7	0.8	34	6 5457090-5	Patent No. 5457090
22	7	0.8	34	6 5495001-16	Patent No. 5495001
23	7	0.8	35	6 5171673-7	Patent No. 5171673
24	7	0.8	69	4 US-09-149-476-601	Sequence 601, App
25	7	0.8	70	4 US-09-149-476-465	Sequence 465, App
26	7	0.8	95	1 US-08-438-753B-24	Sequence 24, Appli
27	7	0.8	95	1 US-08-443-883A-24	Sequence 24, Appli

28	7	0.8	95	2 US-08-631-328-24	Sequence 24, Appli
29	7	0.8	95	2 US-08-455-524B-24	Sequence 24, Appli
30	7	0.8	95	2 US-08-455-021B-24	Sequence 24, Appli
31	7	0.8	159	3 US-09-045-467-24	Sequence 2, Appli
32	7	0.8	175	2 US-08-796-792-2	Sequence 2, Appli
33	7	0.8	175	2 US-08-245-511-26	Sequence 26, Appli
34	7	0.8	175	2 US-08-600-993A-26	Sequence 12, Appli
35	7	0.8	264	4 US-09-323-872A-12	Sequence 12, Appli
36	7	0.8	349	2 US-09-724-864-64	Sequence 64, Appli
37	7	0.8	350	4 US-08-465-971B-2	Sequence 2, Appli
38	7	0.8	355	4 US-09-134-001C-5150	Sequence 5150, Ap
39	7	0.8	355	4 US-09-325-932A-191	Sequence 191, App
40	7	0.8	397	6 5187089-2	Patent No. 5187089
41	7	0.8	397	6 5187089-9	Patent No. 5187089
42	7	0.8	397	6 5187089-10	Patent No. 5187089
43	7	0.8	397	6 5187089-11	Patent No. 5187089
44	7	0.8	397	6 5457090-2	Patent No. 5457090
45	7	0.8	397	6 5457090-4	Patent No. 5457090
46	7	0.8	397	6 5495001-7	Patent No. 5495001
47	7	0.8	398	6 5187089-4	Patent No. 5187089
48	7	0.8	398	6 5187089-5	Patent No. 5187089
49	7	0.8	398	6 5187089-6	Patent No. 5187089
50	7	0.8	398	6 5187089-7	Patent No. 5187089
51	7	0.8	398	6 5495001-9	Patent No. 5495001
52	7	0.8	434	1 US-08-337-602-2	Sequence 2, Appli
53	7	0.8	434	3 US-08-558-135-2	Sequence 2, Appli
54	7	0.8	488	4 US-08-985-343-1	Sequence 1, Appli
55	7	0.8	511	1 US-08-468-700-35	Sequence 35, Appli
56	7	0.8	511	1 US-08-645-971-3	Sequence 3, Appli
57	7	0.8	511	1 US-08-468-220-33	Sequence 33, Appli
58	7	0.8	511	2 US-08-468-698-33	Sequence 33, Appli
59	7	0.8	511	2 US-08-704-706A-35	Sequence 35, Appli
60	7	0.8	511	3 US-08-890-383-4	Sequence 2, Appli
61	7	0.8	511	3 US-08-890-383-4	Sequence 4, Appli
62	7	0.8	511	3 US-08-914-679A-2	Sequence 2, Appli
63	7	0.8	511	3 US-08-914-679A-4	Sequence 4, Appli
64	7	0.8	511	4 US-08-985-659-36	Sequence 36, Appli
65	7	0.8	511	4 US-08-194-664A-33	Sequence 33, Appli
66	7	0.8	511	5 PCT-US94-01553A-33	Sequence 33, Appli
67	7	0.8	511	5 PCT-US95-10426-33	Sequence 33, Appli
68	7	0.8	512	1 US-07-623-953-5	Sequence 5, Appli
69	7	0.8	512	1 US-08-720-899-2	Sequence 2, Appli
70	7	0.8	512	1 US-08-459-610-2	Sequence 2, Appli
71	7	0.8	512	2 US-08-343-804-2	Sequence 2, Appli
72	7	0.8	512	2 US-08-687-399-2	Sequence 2, Appli
73	7	0.8	512	2 US-08-600-908A-2	Sequence 2, Appli
74	7	0.8	512	3 US-08-683-838A-2	Sequence 2, Appli
75	7	0.8	512	4 US-08-985-659-34	Sequence 34, Appli
76	7	0.8	516	4 US-09-636-252A-2	Sequence 2, Appli
77	7	0.8	521	1 US-08-868-373-14	Sequence 14, Appli
78	7	0.8	521	1 US-08-915-003-7	Sequence 7, Appli
79	7	0.8	534	4 US-08-642-247-7	Sequence 7, Appli
80	7	0.8	534	4 US-09-199-677A-67	Sequence 67, Appli
81	7	0.8	551	4 US-09-615-192A-348	Sequence 348, App
82	7	0.8	560	3 US-08-814-052-6	Sequence 6, Appli
83	7	0.8	612	2 US-08-812-829-6	Sequence 6, Appli
84	7	0.8	612	2 US-08-752-307B-11	Sequence 11, Appli
85	7	0.8	612	4 US-09-707-802-11	Sequence 11, Appli
86	7	0.8	631	3 US-09-991-326-11	Sequence 11, Appli
87	7	0.8	631	3 US-08-814-052-8	Sequence 8, Appli
88	7	0.8	631	3 US-08-812-829-8	Sequence 8, Appli
89	7	0.8	655	4 US-09-228-986-70	Sequence 70, Appli
90	7	0.8	719	4 US-09-708-426-4	Sequence 4, Appli
91	7	0.8	750	3 US-08-814-052-2	Sequence 2, Appli
92	7	0.8	750	3 US-08-812-829-2	Sequence 2, Appli
93	7	0.8	1050	3 US-09-045-632-19	Sequence 49, Appli
94	7	0.8	1050	3 US-09-045-632-50	Sequence 50, Appli
95	7	0.8	1068	3 US-08-390-874C-11	Sequence 11, Appli
96	7	0.8	1068	4 US-09-265-772-11	Sequence 11, Appli
97	7	0.8	1069	2 US-08-162-081B-37	Sequence 37, Appli
98	7	0.8	1069	2 US-08-780-872-37	Sequence 37, Appli
99	7	0.8	1069	4 US-09-085-957-37	Sequence 37, Appli
100	7	0.8	1080	2 US-08-162-081B-36	Sequence 36, Appli

101	7	0.8	1080	2	US-08-780-872-36	Sequence 36, Appl
102	7	0.8	1080	4	US-09-085-957-36	Sequence 36, Appl
103	7	0.8	1268	4	US-08-506-2968-28	Sequence 10, Appl
104	7	0.8	2336	1	US-09-268-163-10	Sequence 10, Appl
105	6	0.7	9	1	US-08-787-547-95	Sequence 95, Appl
106	6	0.7	9	3	US-08-159-339A-443	Sequence 443, App
107	6	0.7	9	4	US-09-324-541-1	Sequence 1, Appl1
108	6	0.7	9	4	US-09-357-717-57	Sequence 57, Appl
109	6	0.7	9	4	US-09-305-927-32	Sequence 32, Appl
110	6	0.7	10	4	US-08-564-542A-13	Sequence 13, Appl
111	6	0.7	14	1	US-08-279-0588-3	Sequence 3, Appl1
112	6	0.7	14	1	US-08-828-323-3	Sequence 3, Appl1
113	6	0.7	15	3	US-08-940-095-252	Sequence 252, App
114	6	0.7	15	3	US-08-940-093-252	Sequence 252, App
115	6	0.7	15	3	US-08-940-096-252	Sequence 89, Appl
116	6	0.7	15	4	US-09-248-588-89	Sequence 252, App
117	6	0.7	15	4	US-09-465-719-252	Sequence 252, App
118	6	0.7	15	4	US-09-453-605-252	Sequence 252, App
119	6	0.7	15	4	US-08-279-0588-44	Sequence 44, Appl
120	6	0.7	17	1	US-08-528-129A-8	Sequence 8, Appl1
121	6	0.7	17	2	US-08-828-323-44	Sequence 44, Appl
122	6	0.7	19	4	US-09-146-580-5	Sequence 44, Appl
123	6	0.7	29	2	US-08-318-837-15	Sequence 5, Appl1
124	6	0.7	30	2	US-08-487-890A-41	Sequence 15, Appl
125	6	0.7	30	2	US-08-478-435-41	Sequence 41, Appl
126	6	0.7	30	2	US-08-337-483-41	Sequence 41, Appl
127	6	0.7	30	2	US-08-478-373-41	Sequence 41, Appl
128	6	0.7	30	2	US-08-474-671-41	Sequence 41, Appl
129	6	0.7	30	3	US-08-483-577A-41	Sequence 41, Appl
130	6	0.7	30	3	US-08-897-438-41	Sequence 41, Appl
131	6	0.7	30	4	US-08-637-654-41	Sequence 41, Appl
132	6	0.7	30	4	US-08-649-518-41	Sequence 41, Appl
133	6	0.7	34	2	US-08-468-558-8	Sequence 8, Appl1
134	6	0.7	34	4	US-08-676-444-8	Sequence 8, Appl1
135	6	0.7	34	6	5514582-24	Patent No. 5514582
136	6	0.7	35	4	US-09-012-669F-15	Sequence 15, Appl
137	6	0.7	35	4	US-09-012-669F-16	Sequence 16, Appl
138	6	0.7	35	4	US-09-012-669F-17	Sequence 17, Appl
139	6	0.7	35	4	US-09-012-669F-18	Sequence 18, Appl
140	6	0.7	35	4	US-09-012-669F-22	Sequence 22, Appl
141	6	0.7	40	3	US-09-037-524-4	Sequence 23, Appl
142	6	0.7	40	3	US-09-012-669F-22	Sequence 26, Appl
143	6	0.7	40	4	US-09-012-669F-23	Sequence 27, Appl
144	6	0.7	40	4	US-09-012-669F-26	Sequence 69, Appl
145	6	0.7	40	4	US-09-012-669F-27	Sequence 69, Appl
146	6	0.7	41	1	US-08-468-574B-69	Sequence 1, Appl1
147	6	0.7	41	1	US-08-780-571-69	Sequence 1, Appl1
148	6	0.7	41	3	US-08-888-381-1	Sequence 13, Appl
149	6	0.7	41	4	US-09-012-669F-13	Sequence 21, Appl
150	6	0.7	41	4	US-09-012-669F-21	Sequence 21, Appl
151	6	0.7	41	4	US-09-012-669F-25	Sequence 25, Appl
152	6	0.7	41	4	US-09-012-669F-41	Sequence 99, Appl
153	6	0.7	42	1	US-08-056-200-99	Sequence 5, Appl1
154	6	0.7	42	2	US-08-800-644-99	Sequence 8, Appl1
155	6	0.7	42	4	US-09-012-669F-8	Sequence 3, Appl1
156	6	0.7	42	4	US-09-012-669F-8	Sequence 3, Appl1
157	6	0.7	43	4	US-09-012-669F-12	Sequence 12, Appl
158	6	0.7	43	4	US-09-012-669F-24	Sequence 24, Appl
159	6	0.7	43	4	US-09-012-669F-40	Sequence 40, Appl
160	6	0.7	43	5	PCT-US93-05640-20	Sequence 20, Appl
161	6	0.7	46	3	US-09-053-197A-45	Sequence 45, Appl
162	6	0.7	46	4	US-09-085-761A-50	Sequence 50, Appl
163	6	0.7	46	4	US-09-324-541-2	Sequence 23, Appl
164	6	0.7	59	4	US-09-381-546-23	Sequence 23, Appl
165	6	0.7	59	4	US-09-381-546-25	Sequence 25, Appl
166	6	0.7	62	1	US-07-602-668-12	Sequence 12, Appl
167	6	0.7	62	1	US-07-602-847C-25	Sequence 11, Appl
168	6	0.7	68	1	US-07-965-674-11	Sequence 11, Appl
169	6	0.7	68	1	US-09-134-001C-3656	Sequence 3656, Ap
170	6	0.7	68	5	PCT-US93-09523-11	Sequence 11, Appl
171	6	0.7	71	6	5187079-2	Patent No. 5187079
172	6	0.7	71	6	US-09-100-804-29	Sequence 29, Appl
173	6	0.7	78	3		

174	6	0.7	79	2	US-08-612-788-7	Sequence 7, Appl1
175	6	0.7	79	2	US-08-612-788-11	Sequence 11, Appl
176	6	0.7	79	3	US-09-066-028-7	Sequence 7, Appl1
177	6	0.7	79	3	US-09-066-028-7	Sequence 11, Appl
178	6	0.7	84	4	US-09-382-155-2	Sequence 2, Appl1
179	6	0.7	84	4	US-09-074-044A-2	Sequence 2, Appl1
180	6	0.7	86	4	US-09-012-669F-4	Sequence 4, Appl1
181	6	0.7	91	2	US-08-598-873-19	Sequence 19, Appl
182	6	0.7	91	4	US-08-605-430-19	Sequence 19, Appl
183	6	0.7	95	1	US-07-987-272A-9	Sequence 9, Appl1
184	6	0.7	96	4	US-08-945-983-9	Sequence 9, Appl1
185	6	0.7	98	3	US-09-003-708A-6	Sequence 6, Appl1
186	6	0.7	98	4	US-09-383-212-6	Sequence 6, Appl1
187	6	0.7	100	2	US-08-464-517-10	Sequence 10, Appl
188	6	0.7	100	2	US-08-246-361A-10	Sequence 10, Appl
189	6	0.7	100	3	US-08-463-772-10	Sequence 10, Appl
190	6	0.7	100	5	PCT-US93-05000-10	Sequence 10, Appl
191	6	0.7	103	6	5171684-5	Patent No. 5171684
192	6	0.7	104	6	US-09-146-580-2	Sequence 2, Appl1
193	6	0.7	107	4	5242821-8	Patent No. 5242821
194	6	0.7	109	4	US-08-905-223-313	Sequence 313, App
195	6	0.7	110	4	US-09-134-001C-4768	Sequence 4768, Ap
196	6	0.7	112	2	US-08-997-080-78	Sequence 78, Appl
197	6	0.7	112	2	US-08-997-362-78	Sequence 78, Appl
198	6	0.7	112	3	US-08-873-970-78	Sequence 78, Appl
199	6	0.7	112	4	US-09-095-855-78	Sequence 78, Appl
200	6	0.7	112	4	US-09-324-542-78	Sequence 78, Appl
201	6	0.7	112	4	US-09-205-426-78	Sequence 78, Appl
202	6	0.7	114	2	US-08-844-056-4	Sequence 4, Appl1
203	6	0.7	120	4	US-09-325-932A-84	Sequence 84, Appl
204	6	0.7	127	4	US-09-342-647-12	Sequence 12, Appl
205	6	0.7	130	4	US-09-146-580-1	Sequence 1, Appl1
206	6	0.7	133	4	US-09-325-932A-164	Sequence 164, App
207	6	0.7	133	4	US-09-570-921-34	Sequence 54, Appl
208	6	0.7	134	3	US-09-570-921-32	Sequence 52, Appl
209	6	0.7	140	3	US-08-850-961-12	Sequence 12, Appl
210	6	0.7	140	6	US-09-479-776-12	Sequence 12, Appl
211	6	0.7	141	3	5164490-8	Patent No. 5164490
212	6	0.7	141	3	US-08-850-961-10	Sequence 10, Appl
213	6	0.7	142	4	US-09-479-776-10	Sequence 10, Appl
214	6	0.7	142	4	US-08-945-983-7	Sequence 7, Appl1
215	6	0.7	147	2	US-08-858-207A-518	Sequence 518, App
216	6	0.7	148	2	US-08-460-694-6	Sequence 6, Appl1
217	6	0.7	148	3	US-07-667-711B-6	Sequence 6, Appl1
218	6	0.7	157	4	US-09-518-946-23	Sequence 23, Appl
219	6	0.7	160	2	US-08-612-788-34	Sequence 34, Appl
220	6	0.7	160	2	US-08-612-788-38	Sequence 38, Appl
221	6	0.7	160	3	US-09-066-028-34	Sequence 34, Appl
222	6	0.7	160	3	US-09-066-028-34	Sequence 38, Appl
223	6	0.7	163	5	PCT-US95-08295-5	Sequence 284, App
224	6	0.7	164	4	US-08-858-207A-502	Sequence 502, App
225	6	0.7	169	4	US-09-342-084-6	Sequence 6, Appl1
226	6	0.7	169	4	US-09-516-914-7	Sequence 7, Appl1
227	6	0.7	170	1	5180811-11	Patent No. 5180811
228	6	0.7	171	1	US-08-193-977-5	Sequence 5, Appl1
229	6	0.7	176	4	US-08-969-683A-65	Sequence 65, Appl
230	6	0.7	176	4	US-09-134-001C-3018	Patent No. 5180811
231	6	0.7	176	6	5180811-12	Sequence 3018, Ap
232	6	0.7	182	4	US-08-691-563C-90	Sequence 90, Appl
233	6	0.7	182	4	US-09-522-433B-25	Sequence 25, Appl
234	6	0.7	184	4	US-09-475-316A-99	Sequence 99, Appl
235	6	0.7	184	4	US-08-026-738-23	Sequence 23, Appl
236	6	0.7	186	4	US-09-325-932A-72	Sequence 72, Appl
237	6	0.7	186	6	5180811-13	Patent No. 5180811
238	6	0.7	191	4	US-08-858-207A-274	Sequence 274, App
239	6	0.7	191	4	US-09-397-992A-24	Sequence 24, Appl
240	6	0.7	193	4	US-09-597-576-2	Sequence 2, Appl1
241	6	0.7	193	4	US-08-591-989-4	Sequence 3720, Ap
242	6	0.7	201	4	US-09-134-001C-3720	Sequence 4, Appl1
243	6	0.7	204	1	US-09-053-197A-29	Sequence 29, Appl
244	6	0.7	204	3	US-09-085-761A-29	Sequence 29, Appl
245	6	0.7	204	4		
246	6	0.7	204	4		

247	6	0.7	205	4	US-09-724-864-37	Sequence 37, Appl	320	6	0.7	309	4	US-09-205-426-118	Sequence 118, App
248	6	0.7	207	3	US-08-957-302A-10	Sequence 10, Appl	321	6	0.7	309	4	US-09-651-200-22	Sequence 22, Appl
249	6	0.7	207	4	US-09-542-403-10	Sequence 10, Appl	322	6	0.7	309	5	PCT-US95-02576-21	Sequence 21, Appl
250	6	0.7	210	3	US-08-985-526-21	Sequence 21, Appl	323	6	0.7	311	2	US-08-318-837-9	Sequence 9, Appl
251	6	0.7	214	4	US-09-587-066-6	Sequence 6, Appl	324	6	0.7	312	4	US-08-961-083-12	Sequence 12, Appl
252	6	0.7	214	4	US-08-999-774A-4	Sequence 4, Appl	325	6	0.7	313	3	US-09-362-506-2	Sequence 2, Appl
253	6	0.7	221	4	US-09-382-155-17	Sequence 17, Appl	326	6	0.7	314	4	US-08-205-697A-13	Sequence 13, Appl
254	6	0.7	221	4	US-09-074-044A-17	Sequence 17, Appl	327	6	0.7	314	4	US-08-702-525-13	Sequence 13, Appl
255	6	0.7	222	1	US-08-090-048-1	Sequence 1, Appl	328	6	0.7	314	5	PCT-US95-02576-13	Sequence 13, Appl
256	6	0.7	222	2	US-08-292-550-1	Sequence 1, Appl	329	6	0.7	315	1	US-07-729-099-1	Sequence 1, Appl
257	6	0.7	222	2	US-07-927-661A-1	Sequence 1, Appl	330	6	0.7	315	1	US-08-257-392-1	Sequence 1, Appl
258	6	0.7	223	2	US-08-121-136A-2	Sequence 2, Appl	331	6	0.7	315	3	US-08-770-035-1	Sequence 1, Appl
259	6	0.7	223	4	US-09-254-733-7	Sequence 7, Appl	332	6	0.7	317	1	US-07-688-299-1	Sequence 1, Appl
260	6	0.7	225	2	US-08-886-765-2	Sequence 2, Appl	333	6	0.7	317	1	US-07-980-517A-1	Sequence 1, Appl
261	6	0.7	225	4	US-09-115-660-2	Sequence 2, Appl	334	6	0.7	318	1	US-07-688-299-3	Sequence 3, Appl
262	6	0.7	228	4	US-09-724-864-46	Sequence 46, Appl	335	6	0.7	318	1	US-07-688-299-13	Sequence 13, Appl
263	6	0.7	229	4	US-09-604-978-9	Sequence 9, Appl	336	6	0.7	321	1	US-08-362-670B-26	Sequence 26, Appl
264	6	0.7	242	2	US-08-622-352A-3	Sequence 3, Appl	337	6	0.7	321	1	US-08-333-576C-26	Sequence 26, Appl
265	6	0.7	242	2	US-08-826-390-3	Sequence 3, Appl	338	6	0.7	321	5	US-08-808-324-26	Sequence 26, Appl
266	6	0.7	250	2	US-08-612-788-29	Sequence 29, Appl	339	6	0.7	321	5	PCT-US94-14030A-26	Sequence 26, Appl
267	6	0.7	250	2	US-08-612-788-33	Sequence 33, Appl	340	6	0.7	323	1	US-09-462-844-4	Sequence 4, Appl
268	6	0.7	250	3	US-09-066-028-29	Sequence 29, Appl	341	6	0.7	325	1	US-08-276-919-4	Sequence 4, Appl
269	6	0.7	250	3	US-09-066-028-33	Sequence 33, Appl	342	6	0.7	325	1	US-08-276-919-13	Sequence 13, Appl
270	6	0.7	259	4	US-08-961-083-174	Sequence 174, App	343	6	0.7	325	1	US-08-776-088-4	Sequence 4, Appl
271	6	0.7	259	4	US-09-261-599B-3	Sequence 3, Appl	344	6	0.7	325	1	US-08-776-088-18	Sequence 18, Appl
272	6	0.7	259	4	US-09-456-455A-3	Sequence 3, Appl	345	6	0.7	325	3	US-08-581-148C-4	Sequence 4, Appl
273	6	0.7	266	4	US-09-199-637A-157	Sequence 157, App	346	6	0.7	325	5	PCT-US95-09145A-4	Sequence 4, Appl
274	6	0.7	266	4	US-09-134-001C-5453	Sequence 5453, Ap	347	6	0.7	325	5	PCT-US95-09145A-18	Sequence 18, Appl
275	6	0.7	271	1	US-08-276-919-10	Sequence 10, Appl	348	6	0.7	327	2	US-08-997-080-162	Sequence 162, App
276	6	0.7	271	1	US-08-776-088-13	Sequence 13, Appl	349	6	0.7	327	2	US-08-997-362-162	Sequence 162, App
277	6	0.7	271	5	PCT-US95-09145A-13	Sequence 15, Appl	350	6	0.7	327	4	US-09-095-855-162	Sequence 162, App
278	6	0.7	274	1	US-08-776-088-15	Sequence 15, Appl	351	6	0.7	327	4	US-09-324-542-162	Sequence 162, App
279	6	0.7	274	2	US-08-978-404B-5	Sequence 5, Appl	352	6	0.7	327	4	US-09-134-001C-3511	Sequence 3511, Ap
280	6	0.7	274	5	PCT-US95-09145A-15	Sequence 15, Appl	353	6	0.7	327	4	US-09-205-426-162	Sequence 162, App
281	6	0.7	278	4	US-09-060-410-17	Sequence 17, Appl	354	6	0.7	328	3	US-08-486-099-92	Sequence 92, Appl
282	6	0.7	281	4	US-09-134-001C-3491	Sequence 3491, Ap	355	6	0.7	328	3	US-08-360-107A-102	Sequence 102, App
283	6	0.7	282	1	US-08-324-301-15	Sequence 15, Appl	356	6	0.7	328	3	US-08-484-223B-92	Sequence 92, Appl
284	6	0.7	282	5	PCT-US94-09752-3	Sequence 3, Appl	357	6	0.7	328	3	US-08-919-597-92	Sequence 92, Appl
285	6	0.7	291	1	US-08-358-171-2	Sequence 2, Appl	358	6	0.7	328	3	US-08-475-668A-92	Sequence 92, Appl
286	6	0.7	291	3	US-09-090-947-2	Sequence 2, Appl	359	6	0.7	328	3	US-08-485-551A-92	Sequence 92, Appl
287	6	0.7	291	4	US-09-275-742-2	Sequence 2, Appl	360	6	0.7	328	3	US-08-471-913A-92	Sequence 92, Appl
288	6	0.7	293	4	US-09-134-001C-5333	Sequence 5333, Ap	361	6	0.7	328	3	US-08-485-264A-92	Sequence 92, Appl
289	6	0.7	294	4	US-08-973-334-2	Sequence 2, Appl	362	6	0.7	328	4	US-08-474-349A-92	Sequence 92, Appl
290	6	0.7	294	4	US-09-563-869A-2	Sequence 2, Appl	363	6	0.7	328	4	US-08-255-208A-28	Sequence 28, Appl
291	6	0.7	294	4	US-08-549-489-2	Sequence 2, Appl	364	6	0.7	330	4	US-09-134-001C-3582	Sequence 3582, Ap
292	6	0.7	297	2	US-09-027-013-3	Sequence 3, Appl	365	6	0.7	330	4	US-09-134-001C-3582	Sequence 3582, Ap
293	6	0.7	297	3	US-09-173-581-1	Sequence 1, Appl	366	6	0.7	331	3	US-08-356-419-25	Sequence 25, Appl
294	6	0.7	297	3	US-09-244-233-3	Sequence 3, Appl	367	6	0.7	333	4	US-09-097-319A-2	Sequence 2, Appl
295	6	0.7	297	4	US-08-845-258-26	Sequence 26, Appl	368	6	0.7	334	4	US-08-933-750C-16	Sequence 16, Appl
296	6	0.7	297	4	US-08-990-571-26	Sequence 26, Appl	369	6	0.7	334	4	US-09-234-613-16	Sequence 16, Appl
297	6	0.7	297	4	US-09-420-915-1	Sequence 1, Appl	370	6	0.7	336	4	US-09-457-040B-21	Sequence 21, Appl
298	6	0.7	297	4	US-08-723-142A-26	Sequence 26, Appl	371	6	0.7	337	4	US-09-134-001C-3799	Sequence 3799, Ap
299	6	0.7	297	4	US-09-528-784A-26	Sequence 26, Appl	372	6	0.7	337	4	US-08-248-629A-2	Sequence 2, Appl
300	6	0.7	299	4	US-09-134-001C-5227	Sequence 5227, Ap	373	6	0.7	339	1	US-08-248-629A-6	Sequence 6, Appl
301	6	0.7	300	4	US-09-254-465A-10	Sequence 10, Appl	374	6	0.7	339	1	US-08-451-932-2	Sequence 2, Appl
302	6	0.7	301	4	US-09-288-143-98	Sequence 98, Appl	375	6	0.7	339	1	US-08-451-932-6	Sequence 6, Appl
303	6	0.7	303	4	US-08-158-735A-14	Sequence 14, Appl	376	6	0.7	339	1	US-08-452-260-2	Sequence 2, Appl
304	6	0.7	303	4	US-08-818-112-92	Sequence 92, Appl	377	6	0.7	339	1	US-08-452-260-6	Sequence 6, Appl
305	6	0.7	303	4	US-08-818-111-93	Sequence 93, Appl	378	6	0.7	339	1	US-08-326-785-2	Sequence 2, Appl
306	6	0.7	303	4	US-09-056-556-92	Sequence 92, Appl	379	6	0.7	339	1	US-08-326-785-6	Sequence 6, Appl
307	6	0.7	303	4	US-09-651-200-23	Sequence 23, Appl	380	6	0.7	339	2	US-08-612-788-2	Sequence 2, Appl
308	6	0.7	303	4	US-09-072-596-93	Sequence 93, Appl	381	6	0.7	339	2	US-08-612-788-6	Sequence 6, Appl
309	6	0.7	304	4	US-09-134-001C-3492	Sequence 3492, Ap	382	6	0.7	339	2	US-08-605-598B-2	Sequence 2, Appl
310	6	0.7	306	4	US-09-342-647-26	Sequence 26, Appl	383	6	0.7	339	2	US-08-605-598B-6	Sequence 6, Appl
311	6	0.7	309	2	US-08-456-104-4	Sequence 4, Appl	384	6	0.7	339	2	US-08-429-743-2	Sequence 2, Appl
312	6	0.7	309	2	US-08-997-080-118	Sequence 118, App	385	6	0.7	339	2	US-08-429-743-6	Sequence 6, Appl
313	6	0.7	309	3	US-08-997-362-118	Sequence 118, App	386	6	0.7	339	2	US-08-866-735-2	Sequence 2, Appl
314	6	0.7	309	3	US-08-479-744A-23	Sequence 23, Appl	387	6	0.7	339	2	US-08-866-735-6	Sequence 6, Appl
315	6	0.7	309	4	US-08-280-757B-23	Sequence 23, Appl	388	6	0.7	339	3	US-09-066-028-2	Sequence 2, Appl
316	6	0.7	309	4	US-09-095-855-118	Sequence 118, App	389	6	0.7	339	3	US-09-066-028-6	Sequence 6, Appl
317	6	0.7	309	4	US-08-205-697A-21	Sequence 21, Appl	390	6	0.7	339	5	PCT-US95-05107-2	Sequence 2, Appl
318	6	0.7	309	4	US-08-702-525-21	Sequence 21, Appl	391	6	0.7	339	5	PCT-US95-05107-6	Sequence 6, Appl
319	6	0.7	309	4	US-09-324-542-118	Sequence 118, App	392	6	0.7	341	1	US-08-396-957A-5	Sequence 5, Appl

393	6	0.7	341	3	US-08-389-812-4	Sequence 4, Appl	466	6	0.7	383	2	US-08-760-936-4	Sequence 4, Appl
394	6	0.7	341	3	US-09-270-813-4	Sequence 4, Appl	467	6	0.7	384	2	US-08-935-887-1	Sequence 1, Appl
395	6	0.7	341	3	US-08-591-605-6	Sequence 6, Appl	468	6	0.7	385	1	US-08-758-213-1	Sequence 1, Appl
396	6	0.7	343	4	US-08-279-058B-9	Sequence 9, Appl	469	6	0.7	386	2	US-08-692-787-48	Sequence 11, Appl
397	6	0.7	343	4	US-08-828-323-9	Sequence 9, Appl	470	6	0.7	386	4	US-09-176-666-11	Sequence 48, Appl
398	6	0.7	344	4	US-09-134-001C-5004	Sequence 5004, Ap	471	6	0.7	386	4	US-09-176-666-5	Sequence 10, Appl
399	6	0.7	345	4	US-09-364-230-8	Sequence 8, Appl	472	6	0.7	387	4	US-09-176-666-10	Sequence 9, Appl
400	6	0.7	346	2	US-08-702-153-2	Sequence 2, Appl	473	6	0.7	388	4	US-09-176-666-9	Sequence 8, Appl
401	6	0.7	346	2	US-08-802-153-4	Sequence 4, Appl	474	6	0.7	389	4	US-09-176-666-8	Sequence 3161, Ap
402	6	0.7	346	2	US-08-702-153-2	Sequence 16, Appl	475	6	0.7	389	4	US-09-134-001C-3161	Sequence 3161, Ap
403	6	0.7	351	1	US-08-468-847B-16	Sequence 68, Appl	476	6	0.7	390	4	US-09-176-666-7	Sequence 6, Appl
404	6	0.7	352	1	US-08-196-989B-16	Sequence 2, Appl	477	6	0.7	391	4	US-09-176-666-6	Sequence 7, Appl
405	6	0.7	352	2	US-08-612-788-39	Sequence 39, Appl	478	6	0.7	391	4	US-09-123-912-110	Sequence 110, App
406	6	0.7	352	2	US-08-760-936-2	Sequence 2, Appl	479	6	0.7	391	4	US-09-643-597-110	Sequence 6, Appl
407	6	0.7	352	3	US-09-066-028-39	Sequence 39, Appl	480	6	0.7	392	2	US-08-387-942C-23	Sequence 110, App
408	6	0.7	354	1	US-08-276-919-2	Sequence 2, Appl	481	6	0.7	392	4	US-08-176-666-5	Sequence 23, Appl
409	6	0.7	354	1	US-08-776-088-2	Sequence 6, Appl	482	6	0.7	393	4	US-09-176-666-4	Sequence 5, Appl
410	6	0.7	354	1	US-08-776-088-2	Sequence 2, Appl	483	6	0.7	393	6	512669-6	Sequence 4, Appl
411	6	0.7	354	3	US-09-325-320-2	Sequence 2, Appl	484	6	0.7	393	6	512669-6	Sequence 4, Appl
412	6	0.7	354	3	US-09-585-109-2	Sequence 2, Appl	485	6	0.7	394	4	US-09-176-666-3	Sequence 3, Appl
413	6	0.7	354	5	PCT-US95-09145A-2	Sequence 2, Appl	486	6	0.7	395	1	US-07-931-943-2	Sequence 2, Appl
414	6	0.7	355	4	PCT-US95-09145A-6	Sequence 6, Appl	487	6	0.7	395	2	US-08-624-601-15	Sequence 15, Appl
415	6	0.7	355	1	US-08-153-848-28	Sequence 28, Appl	488	6	0.7	396	4	US-08-948-997-5	Sequence 2, Appl
416	6	0.7	355	1	US-08-153-848-32	Sequence 32, Appl	489	6	0.7	396	4	US-08-948-997-5	Sequence 5, Appl
417	6	0.7	355	3	US-09-299-843A-28	Sequence 28, Appl	490	6	0.7	397	3	US-09-348-817A-5	Sequence 15, Appl
418	6	0.7	355	4	US-09-299-843A-32	Sequence 32, Appl	491	6	0.7	397	4	US-09-463-238-15	Sequence 26, Appl
419	6	0.7	355	4	US-09-088-337B-28	Sequence 28, Appl	492	6	0.7	398	3	US-09-189-035-6	Sequence 6, Appl
420	6	0.7	355	4	US-09-088-337B-32	Sequence 32, Appl	493	6	0.7	398	4	US-08-928-881-26	Sequence 26, Appl
421	6	0.7	355	4	US-09-134-001C-3580	Sequence 3580, Ap	494	6	0.7	398	4	US-09-382-086-6	Sequence 6, Appl
422	6	0.7	355	5	PCT-US93-11153-28	Sequence 28, Appl	495	6	0.7	398	4	US-09-176-666-1	Sequence 1, Appl
423	6	0.7	355	5	PCT-US93-11153-32	Sequence 32, Appl	496	6	0.7	398	4	US-09-543-921-26	Sequence 26, Appl
424	6	0.7	357	1	US-08-468-847B-17	Sequence 17, Appl	497	6	0.7	398	4	US-09-266-014-4	Sequence 4, Appl
425	6	0.7	357	1	US-09-253-316-25	Sequence 25, Appl	498	6	0.7	399	4	US-08-765-907A-10	Sequence 10, Appl
426	6	0.7	358	1	US-08-034-650-10	Sequence 10, Appl	499	6	0.7	399	4	US-09-134-001C-4571	Sequence 4571, Ap
427	6	0.7	358	1	US-08-700-186-2	Sequence 2, Appl	500	6	0.7	400	1	US-07-730-953-2	Sequence 2, Appl
428	6	0.7	358	2	US-08-449-015-10	Sequence 10, Appl	501	6	0.7	400	1	US-08-961-083-190	Sequence 190, App
429	6	0.7	358	2	US-08-558-823-19	Sequence 19, Appl	502	6	0.7	400	4	US-09-123-912-112	Sequence 112, App
430	6	0.7	358	2	US-08-914-981-2	Sequence 2, Appl	503	6	0.7	400	4	US-09-643-597-112	Sequence 10, Appl
431	6	0.7	358	3	US-09-116-115-2	Sequence 11, Appl	504	6	0.7	402	4	US-09-029-755C-5	Sequence 5, Appl
432	6	0.7	358	4	US-09-541-762-2	Sequence 2, Appl	505	6	0.7	402	4	US-08-533-669A-10	Sequence 2, Appl
433	6	0.7	359	4	US-08-809-103B-2	Sequence 2, Appl	506	6	0.7	403	2	US-08-607-509-4	Sequence 2, Appl
434	6	0.7	359	4	US-08-809-103B-4	Sequence 4, Appl	507	6	0.7	403	2	US-08-607-509-2	Sequence 2, Appl
435	6	0.7	359	4	US-08-809-103B-6	Sequence 6, Appl	508	6	0.7	403	2	US-08-607-509-4	Sequence 2, Appl
436	6	0.7	359	4	US-08-809-103B-8	Sequence 8, Appl	509	6	0.7	403	2	US-08-607-509-4	Sequence 2, Appl
437	6	0.7	362	4	US-09-625-188-8	Sequence 8, Appl	510	6	0.7	403	2	US-08-634-642-2	Sequence 2, Appl
438	6	0.7	362	4	US-09-134-001C-5403	Sequence 5403, Ap	511	6	0.7	403	2	US-08-634-642-4	Sequence 4, Appl
439	6	0.7	363	1	US-08-022-411-2	Sequence 2, Appl	512	6	0.7	403	3	US-08-989-370-2	Sequence 3, Appl
440	6	0.7	363	1	US-08-984-618-13	Sequence 13, Appl	513	6	0.7	403	3	US-08-989-370-2	Sequence 4, Appl
441	6	0.7	368	1	US-08-093-372-4	Sequence 4, Appl	514	6	0.7	403	4	US-09-183-861-10	Sequence 10, Appl
442	6	0.7	368	1	US-08-093-372-4	Sequence 4, Appl	515	6	0.7	403	4	US-09-022-765-10	Sequence 10, Appl
443	6	0.7	368	4	US-09-000-092-2	Sequence 4, Appl	516	6	0.7	403	5	PCT-US95-05064-2	Sequence 2, Appl
444	6	0.7	368	4	US-09-000-092-4	Sequence 6, Appl	517	6	0.7	403	5	PCT-US95-05064-2	Sequence 2, Appl
445	6	0.7	368	4	US-09-000-092-6	Sequence 8, Appl	518	6	0.7	405	2	US-08-519-081-2	Sequence 2, Appl
446	6	0.7	368	4	US-09-000-092-8	Sequence 10, Appl	519	6	0.7	405	2	US-08-754-356C-2	Sequence 2, Appl
447	6	0.7	368	4	US-09-000-092-10	Sequence 14, Appl	520	6	0.7	409	4	US-09-029-755C-2	Sequence 2, Appl
448	6	0.7	374	1	US-09-413-574-14	Sequence 14, Appl	521	6	0.7	411	1	US-08-399-561-2	Sequence 2, Appl
449	6	0.7	374	1	US-08-096-623A-14	Sequence 14, Appl	522	6	0.7	412	4	US-09-445-472-1	Sequence 4, Appl
450	6	0.7	375	1	US-08-121-714-2	Sequence 2, Appl	523	6	0.7	414	2	US-08-305-764C-58	Sequence 5528, Ap
451	6	0.7	375	1	US-08-736-361A-5	Sequence 5, Appl	524	6	0.7	418	2	US-08-305-764C-58	Sequence 58, Appl
452	6	0.7	375	1	US-08-736-361A-5	Sequence 2, Appl	525	6	0.7	418	2	US-08-305-764C-58	Sequence 60, Appl
453	6	0.7	375	2	US-08-477-108A-2	Sequence 2, Appl	526	6	0.7	422	4	US-09-427-501-2	Sequence 2, Appl
454	6	0.7	375	5	PCT-US93-08322-2	Sequence 41, Appl	527	6	0.7	423	4	US-08-985-526-23	Sequence 23, Appl
455	6	0.7	378	2	US-08-612-788-41	Sequence 41, Appl	528	6	0.7	428	6	5179023-4	Sequence 23, Appl
456	6	0.7	381	2	US-09-066-028-41	Sequence 3, Appl	529	6	0.7	429	1	US-07-602-824A-2	Sequence 2, Appl
457	6	0.7	381	2	US-08-845-566-3	Sequence 28, Appl	530	6	0.7	429	1	US-07-602-824A-2	Sequence 3, Appl
458	6	0.7	381	2	US-08-467-948A-28	Sequence 28, Appl	531	6	0.7	429	1	US-07-983-451-2	Sequence 3, Appl
459	6	0.7	381	3	US-08-852-824-18	Sequence 12, Appl	532	6	0.7	429	1	US-08-261-578-3	Sequence 2, Appl
460	6	0.7	381	3	US-08-467-947A-28	Sequence 12, Appl	533	6	0.7	429	1	US-08-261-577-7	Sequence 7, Appl
461	6	0.7	381	5	PCT-US96-10618-4	Sequence 4, Appl	534	6	0.7	429	1	US-08-261-577-9	Sequence 9, Appl
462	6	0.7	381	5	US-09-542-733-2	Sequence 2, Appl	535	6	0.7	438	4	US-09-070-356-4	Sequence 4, Appl
463	6	0.7	382	4	US-09-262-477-2	Sequence 2, Appl	536	6	0.7	438	5	US-09-134-001C-3139	Sequence 3139, Ap
464	6	0.7	382	4	US-09-262-477-2	Sequence 2, Appl	537	6	0.7	438	5	PCT-US95-05922A-2	Sequence 2, Appl
465	6	0.7	383	1	US-08-196-989B-4	Sequence 4, Appl	538	6	0.7	439	3	US-08-993-359-24	Sequence 24, Appl

539	6	0.7	439	3	US-09-221-654-2	Sequence 2, Appli	612	6	0.7	522	4	US-08-894-818B-3	Sequence 3, Appli
540	6	0.7	439	6	US-08-989-358A-2	Sequence 2, Appli	613	6	0.7	522	4	US-09-445-472-4	Sequence 4, Appli
541	6	0.7	441	2	US-08-713-815A-3	Sequence 3, Appli	614	6	0.7	523	2	US-08-997-080-114	Sequence 114, App
542	6	0.7	442	3	US-08-993-359-26	Sequence 26, Appli	615	6	0.7	523	2	US-08-482-728A-19	Sequence 19, Appli
543	6	0.7	445	3	US-08-859-167-2	Sequence 2, Appli	616	6	0.7	523	2	US-08-923-772-2	Sequence 2, Appli
544	6	0.7	445	3	US-09-109-273-2	Sequence 2, Appli	617	6	0.7	523	2	US-08-997-362-114	Sequence 114, App
545	6	0.7	445	3	US-09-276-993-2	Sequence 2, Appli	618	6	0.7	523	4	US-09-028-366-4	Sequence 4, Appli
546	6	0.7	448	2	US-08-884-072-1	Sequence 1, Appli	619	6	0.7	523	4	US-09-095-855-114	Sequence 114, App
547	6	0.7	448	4	US-09-212-168-1	Sequence 1, Appli	620	6	0.7	523	4	US-09-324-542-114	Sequence 114, App
548	6	0.7	452	1	US-08-191-337-3	Sequence 3, Appli	621	6	0.7	523	4	US-09-385-287-2	Sequence 2, Appli
549	6	0.7	454	4	US-09-323-872A-15	Sequence 15, Appli	622	6	0.7	523	4	US-09-205-426-114	Sequence 114, App
550	6	0.7	452	4	US-09-518-046-2	Sequence 2, Appli	623	6	0.7	524	3	US-08-957-302A-2	Sequence 2, Appli
551	6	0.7	455	4	US-08-870-827-3	Sequence 3, Appli	624	6	0.7	524	4	US-09-299-662-1	Sequence 1, Appli
552	6	0.7	455	4	US-09-308-003-13	Sequence 13, Appli	625	6	0.7	524	4	US-09-542-403-2	Sequence 2, Appli
553	6	0.7	455	4	US-09-317-179-3	Sequence 3, Appli	626	6	0.7	525	1	US-08-681-129-2	Sequence 2, Appli
554	6	0.7	456	1	US-08-021-601-6	Sequence 6, Appli	627	6	0.7	526	2	US-08-541-033A-2	Sequence 2, Appli
555	6	0.7	456	1	US-08-082-849B-6	Sequence 6, Appli	628	6	0.7	526	2	US-08-828-451-2	Sequence 2, Appli
556	6	0.7	456	5	PCT-US94-01624-6	Sequence 6, Appli	629	6	0.7	530	4	US-08-793-044-3	Sequence 3, Appli
557	6	0.7	458	4	US-09-457-046B-61	Sequence 61, Appli	630	6	0.7	534	3	US-08-875-223-8	Sequence 8, Appli
558	6	0.7	459	4	US-09-305-984-20	Sequence 20, Appli	631	6	0.7	538	4	US-09-370-368-9	Sequence 9, Appli
559	6	0.7	459	4	US-09-073-541A-20	Sequence 20, Appli	632	6	0.7	538	4	US-09-309-572-12	Sequence 12, Appli
560	6	0.7	462	1	US-08-417-330A-16	Sequence 16, Appli	633	6	0.7	539	3	US-08-941-445A-21	Sequence 21, Appli
561	6	0.7	465	2	US-08-878-989-18	Sequence 18, Appli	634	6	0.7	540	4	US-09-011-074-4	Sequence 4, Appli
562	6	0.7	465	2	US-08-860-150-7	Sequence 7, Appli	635	6	0.7	540	4	US-08-461-722-4	Sequence 4, Appli
563	6	0.7	465	3	US-09-338-132-7	Sequence 7, Appli	636	6	0.7	540	4	US-08-336-251-4	Sequence 4, Appli
564	6	0.7	465	4	US-09-272-796-18	Sequence 18, Appli	637	6	0.7	540	5	PCT-US94-06362-4	Sequence 4, Appli
565	6	0.7	465	4	US-09-134-001C-3856	Sequence 3856, Ap	638	6	0.7	541	2	US-08-997-080-160	Sequence 160, App
566	6	0.7	472	1	US-08-021-601-8	Sequence 8, Appli	639	6	0.7	541	2	US-08-997-362-160	Sequence 160, App
567	6	0.7	472	1	US-08-082-849B-8	Sequence 8, Appli	640	6	0.7	541	2	US-09-095-855-160	Sequence 160, App
568	6	0.7	472	5	PCT-US94-01624-8	Sequence 8, Appli	641	6	0.7	541	4	US-09-324-542-160	Sequence 160, App
569	6	0.7	479	4	US-09-177-349-3	Sequence 3, Appli	642	6	0.7	541	4	US-09-205-426-160	Sequence 160, App
570	6	0.7	480	4	US-08-795-088A-2	Sequence 2, Appli	643	6	0.7	542	4	US-09-071-035-62	Sequence 62, Appli
571	6	0.7	480	4	US-09-069-023-34	Sequence 34, Appli	644	6	0.7	545	2	US-08-467-822-30	Sequence 30, Appli
572	6	0.7	482	1	US-07-792-885A-1	Sequence 1, Appli	645	6	0.7	545	4	US-08-432-697-10	Sequence 10, Appli
573	6	0.7	482	1	US-08-142-439A-7	Sequence 7, Appli	646	6	0.7	545	4	US-08-466-248-30	Sequence 30, Appli
574	6	0.7	482	2	US-08-869-477-7	Sequence 7, Appli	647	6	0.7	545	3	US-08-470-260-6	Sequence 6, Appli
575	6	0.7	482	4	US-09-155-183-4	Sequence 4, Appli	648	6	0.7	546	3	US-08-471-491-6	Sequence 6, Appli
576	6	0.7	488	4	US-09-039-053-2	Sequence 2, Appli	649	6	0.7	546	4	US-08-466-662-6	Sequence 6, Appli
577	6	0.7	488	4	US-08-939-309-10	Sequence 10, Appli	650	6	0.7	546	4	US-08-256-847C-1	Sequence 1, Appli
578	6	0.7	493	4	US-08-999-774A-12	Sequence 12, Appli	651	6	0.7	546	4	US-08-256-847C-7	Sequence 7, Appli
579	6	0.7	501	2	US-08-980-060-6	Sequence 6, Appli	652	6	0.7	547	4	US-09-312-183A-2	Sequence 2, Appli
580	6	0.7	501	4	US-09-307-185-6	Sequence 6, Appli	653	6	0.7	550	4	US-09-330-245A-2	Sequence 2, Appli
581	6	0.7	503	4	US-08-740-223A-11	Sequence 11, Appli	654	6	0.7	553	2	US-08-387-942C-2	Sequence 2, Appli
582	6	0.7	503	4	US-09-709-188-11	Sequence 11, Appli	655	6	0.7	557	4	US-09-518-386B-1	Sequence 1, Appli
583	6	0.7	506	4	US-09-299-662-4	Sequence 4, Appli	656	6	0.7	557	4	US-09-518-386B-3	Sequence 3, Appli
584	6	0.7	508	1	US-08-021-601-10	Sequence 10, Appli	657	6	0.7	567	4	US-09-347-878-42	Sequence 42, Appli
585	6	0.7	508	1	US-08-082-849B-10	Sequence 10, Appli	658	6	0.7	568	1	US-08-262-424-7	Sequence 7, Appli
586	6	0.7	508	5	PCT-US94-01624-10	Sequence 10, Appli	659	6	0.7	568	2	US-08-493-197-7	Sequence 7, Appli
587	6	0.7	509	1	US-08-149-105-14	Sequence 14, Appli	660	6	0.7	568	2	US-08-717-567A-4	Sequence 4, Appli
588	6	0.7	509	1	US-08-317-847-14	Sequence 14, Appli	661	6	0.7	568	3	US-08-883-610A-4	Sequence 4, Appli
589	6	0.7	509	2	US-08-665-926-8	Sequence 8, Appli	662	6	0.7	568	4	US-09-238-373-2	Sequence 2, Appli
590	6	0.7	509	2	US-08-481-337A-4	Sequence 4, Appli	663	6	0.7	568	4	US-08-936-094A-4	Sequence 4, Appli
591	6	0.7	509	4	US-09-382-256-4	Sequence 4, Appli	664	6	0.7	568	4	US-08-939-309-4	Sequence 4, Appli
592	6	0.7	509	4	US-08-740-223A-10	Sequence 10, Appli	665	6	0.7	568	5	PCT-US95-07844-7	Sequence 7, Appli
593	6	0.7	509	4	US-09-395-115-4	Sequence 4, Appli	666	6	0.7	573	1	US-08-200-512-2	Sequence 2, Appli
594	6	0.7	509	4	US-08-123-934A-6	Sequence 6, Appli	667	6	0.7	577	4	US-09-261-855-2	Sequence 2, Appli
595	6	0.7	509	4	US-08-436-265-4	Sequence 4, Appli	668	6	0.7	579	4	US-09-171-710-6	Sequence 6, Appli
596	6	0.7	509	4	US-09-679-187-4	Sequence 4, Appli	669	6	0.7	579	4	US-09-171-710-6	Sequence 6, Appli
597	6	0.7	509	4	US-09-202-491-2	Sequence 2, Appli	670	6	0.7	583	3	US-08-941-445A-13	Sequence 13, Appli
598	6	0.7	509	4	US-09-202-491-3	Sequence 3, Appli	671	6	0.7	585	4	US-09-173-300-7	Sequence 7, Appli
599	6	0.7	509	5	US-09-709-188-10	Sequence 10, Appli	672	6	0.7	589	1	US-07-668-648-6	Sequence 6, Appli
600	6	0.7	509	5	PCT-US94-10080-6	Sequence 6, Appli	673	6	0.7	589	2	US-08-724-394A-1	Sequence 1, Appli
601	6	0.7	509	5	PCT-US94-11328A-2	Sequence 2, Appli	674	6	0.7	589	2	US-08-429-998-6	Sequence 6, Appli
602	6	0.7	512	1	US-07-623-953-3	Sequence 3, Appli	675	6	0.7	589	2	US-08-431-333-6	Sequence 6, Appli
603	6	0.7	512	2	US-08-541-033A-4	Sequence 4, Appli	676	6	0.7	589	4	US-08-991-862-2	Sequence 2, Appli
604	6	0.7	512	2	US-08-828-451-4	Sequence 4, Appli	677	6	0.7	589	5	PCT-US91-02321-6	Sequence 6, Appli
605	6	0.7	513	3	US-08-948-564-2	Sequence 2, Appli	678	6	0.7	590	2	US-08-736-317-12	Sequence 12, Appli
606	6	0.7	513	3	US-09-146-283-2	Sequence 2, Appli	679	6	0.7	590	2	US-09-134-001C-4515	Sequence 4515, Ap
607	6	0.7	515	3	US-08-579-823A-2	Sequence 2, Appli	680	6	0.7	592	2	US-08-599-171A-30	Sequence 30, Appli
608	6	0.7	515	3	US-08-344-195-2	Sequence 2, Appli	681	6	0.7	592	2	US-08-646-590B-30	Sequence 30, Appli
609	6	0.7	515	4	US-09-312-183A-3	Sequence 3, Appli	682	6	0.7	592	3	US-09-069-226-30	Sequence 30, Appli
610	6	0.7	521	4	US-09-071-035-64	Sequence 64, Appli	683	6	0.7	592	4	US-08-412-184-30	Sequence 30, Appli
611	6	0.7	521	4	US-09-071-035-64	Sequence 64, Appli	684	6	0.7	594	4	US-08-987-151-2	Sequence 2, Appli

685	6	0.7	596	4	US-09-134-001C-3363	Sequence 3363, Ap
686	6	0.7	597	6	5171850-2	Patent No. 5171850
687	6	0.7	599	1	US-07-879-617A-10	Sequence 10, Appl
688	6	0.7	599	1	US-08-295-814E-11	Sequence 11, Appl
689	6	0.7	599	1	US-08-240-783B-4	Sequence 4, Appl
690	6	0.7	599	1	US-08-753-985-10	Sequence 10, Appl
691	6	0.7	599	3	US-09-084-813-4	Sequence 4, Appl
692	6	0.7	599	3	US-09-343-361-11	Sequence 11, Appl
693	6	0.7	599	5	PCT-US92-09662-4	Sequence 4, Appl
694	6	0.7	600	3	US-09-212-971-12	Sequence 12, Appl
695	6	0.7	600	4	US-08-800-929A-12	Sequence 12, Appl
696	6	0.7	600	4	US-09-617-053A-12	Sequence 12, Appl
697	6	0.7	604	2	US-08-511-485-6	Sequence 6, Appl
698	6	0.7	604	3	US-09-212-971-6	Sequence 6, Appl
699	6	0.7	604	3	US-08-800-929A-6	Sequence 6, Appl
700	6	0.7	604	4	US-08-569-749-4	Sequence 4, Appl
701	6	0.7	604	4	US-09-617-053A-6	Sequence 6, Appl
702	6	0.7	604	5	PCT-US96-12860-4	Sequence 4, Appl
703	6	0.7	607	4	US-08-969-683A-59	Sequence 59, Appl
704	6	0.7	607	4	US-09-134-001C-2994	Sequence 2994, Ap
705	6	0.7	609	4	US-09-115-475-16	Sequence 16, Appl
706	6	0.7	609	4	US-09-115-475-19	Sequence 19, Appl
707	6	0.7	609	4	US-09-115-475-22	Sequence 22, Appl
708	6	0.7	609	4	US-09-115-475-25	Sequence 25, Appl
709	6	0.7	609	4	US-09-115-475-31	Sequence 31, Appl
710	6	0.7	612	3	US-09-212-971-14	Sequence 14, Appl
711	6	0.7	612	3	US-08-800-929A-14	Sequence 14, Appl
712	6	0.7	612	4	US-08-569-749-14	Sequence 14, Appl
713	6	0.7	612	4	US-09-617-053A-14	Sequence 14, Appl
714	6	0.7	612	4	US-09-617-053A-14	Sequence 14, Appl
715	6	0.7	612	5	PCT-US96-12860-14	Sequence 14, Appl
716	6	0.7	615	2	US-08-911-445-2	Sequence 2, Appl
717	6	0.7	615	4	US-09-182-983-2	Sequence 2, Appl
718	6	0.7	616	2	US-08-545-745-2	Sequence 2, Appl
719	6	0.7	618	3	US-09-212-971-8	Sequence 8, Appl
720	6	0.7	618	3	US-08-511-485-8	Sequence 8, Appl
721	6	0.7	618	4	US-08-800-929A-8	Sequence 8, Appl
722	6	0.7	618	4	US-08-569-749-2	Sequence 2, Appl
723	6	0.7	618	4	US-09-617-053A-8	Sequence 8, Appl
724	6	0.7	618	4	US-09-617-053A-8	Sequence 8, Appl
725	6	0.7	618	4	US-09-069-023-29	Sequence 29, Appl
726	6	0.7	618	5	US-09-134-001C-4039	Sequence 4039, Ap
727	6	0.7	618	5	PCT-US96-12860-2	Sequence 2, Appl
728	6	0.7	621	1	US-08-134-001C-4248	Sequence 4248, Ap
729	6	0.7	621	4	US-08-295-814E-6	Sequence 6, Appl
730	6	0.7	621	5	PCT-US93-01959-6	Sequence 6, Appl
731	6	0.7	623	3	US-09-104-068-4	Sequence 4, Appl
732	6	0.7	623	3	US-08-581-148C-18	Sequence 18, Appl
733	6	0.7	625	4	US-08-759-436-3	Sequence 4, Appl
734	6	0.7	625	4	US-09-104-068-2	Sequence 2, Appl
735	6	0.7	637	4	US-08-759-436-5	Sequence 5, Appl
736	6	0.7	637	4	US-09-104-068-2	Sequence 2, Appl
737	6	0.7	638	2	US-08-897-443-1	Sequence 1, Appl
738	6	0.7	640	3	US-09-026-343-2	Sequence 3, Appl
739	6	0.7	640	4	US-09-362-871-2	Sequence 2, Appl
740	6	0.7	643	2	US-08-797-358B-3	Sequence 3, Appl
741	6	0.7	647	2	US-08-305-764C-56	Sequence 56, Appl
742	6	0.7	647	2	US-08-844-909-17	Sequence 2, Appl
743	6	0.7	647	2	US-09-300-909-17	Sequence 17, Appl
744	6	0.7	649	4	US-09-192-909-2	Sequence 2, Appl
745	6	0.7	650	3	US-08-362-525-2	Sequence 2, Appl
746	6	0.7	652	4	US-08-987-123-2	Sequence 2, Appl
747	6	0.7	654	4	US-09-061-764A-2	Sequence 2, Appl
748	6	0.7	654	4	US-08-894-818B-35	Sequence 35, Appl
749	6	0.7	654	4	US-09-445-472-16	Sequence 16, Appl
750	6	0.7	659	4	US-08-894-818B-5	Sequence 5, Appl
751	6	0.7	659	4	US-09-228-986-75	Sequence 75, Appl
752	6	0.7	675	3	US-08-776-265-5	Sequence 5, Appl
753	6	0.7	676	4	US-08-947-965-76	Sequence 76, Appl
754	6	0.7	676	4	US-08-760-615-2	Sequence 2, Appl
755	6	0.7	676	4	US-08-693-147-6	Sequence 6, Appl
756	6	0.7	679	4	US-09-177-650-4	Sequence 4, Appl
757	6	0.7	679	4	US-08-913-942-15	Sequence 15, Appl
	6	0.7	679	4	US-09-268-347-26	Sequence 26, Appl
758	6	0.7	682	2	US-08-911-445-3	Sequence 3, Appl
759	6	0.7	682	4	US-09-182-983-3	Sequence 4, Appl
760	6	0.7	686	4	US-09-061-764A-15	Sequence 15, Appl
761	6	0.7	688	4	US-09-071-035-46A	Sequence 46A, App
762	6	0.7	693	4	US-08-235-836C-68	Sequence 68, Appl
763	6	0.7	693	4	US-08-235-836C-72	Sequence 72, Appl
764	6	0.7	693	4	US-08-960-048-11	Sequence 11, Appl
765	6	0.7	703	3	US-08-941-445A-11	Sequence 4, Appl
766	6	0.7	703	3	US-08-910-925-4	Sequence 7, Appl
767	6	0.7	703	3	US-08-425-843-7	Sequence 321, App
768	6	0.7	715	3	US-09-620-412C-321	Sequence 2, Appl
769	6	0.7	725	3	US-08-425-843-2	Sequence 20, Appl
770	6	0.7	729	4	US-09-625-188-20	Sequence 5, Appl
771	6	0.7	745	1	US-08-453-472-5	Sequence 9, Appl
772	6	0.7	745	1	US-08-038-948-9	Sequence 5, Appl
773	6	0.7	745	1	US-08-455-952-5	Sequence 5, Appl
774	6	0.7	745	2	US-08-484-933B-43	Sequence 43, Appl
775	6	0.7	745	2	US-08-862-903-5	Sequence 43, Appl
776	6	0.7	745	2	US-08-862-903-5	Sequence 43, Appl
777	6	0.7	745	2	US-08-484-158B-43	Sequence 43, Appl
778	6	0.7	745	2	US-08-484-158B-43	Sequence 43, Appl
779	6	0.7	745	2	US-08-480-150A-43	Sequence 43, Appl
780	6	0.7	745	3	US-08-458-731-43	Sequence 43, Appl
781	6	0.7	745	3	US-08-149-223A-43	Sequence 43, Appl
782	6	0.7	746	5	PCT-US95-10509-2	Sequence 2, Appl
783	6	0.7	747	3	US-08-816-633A-51	Sequence 51, Appl
784	6	0.7	747	3	US-08-885-291-51	Sequence 51, Appl
785	6	0.7	747	4	US-09-496-672-51	Sequence 51, Appl
786	6	0.7	749	4	US-09-562-737-84	Sequence 84, Appl
787	6	0.7	754	4	US-09-521-780-2	Sequence 2, Appl
788	6	0.7	754	4	US-09-521-780-3	Sequence 3, Appl
789	6	0.7	766	4	US-09-463-238-6	Sequence 6, Appl
790	6	0.7	769	3	US-09-134-078-12	Sequence 12, Appl
791	6	0.7	772	4	US-09-320-878-12	Sequence 12, Appl
792	6	0.7	776	1	US-08-021-601-2	Sequence 2, Appl
793	6	0.7	776	5	PCT-US94-01624-2	Sequence 2, Appl
794	6	0.7	776	5	US-08-907-166-6	Sequence 6, Appl
795	6	0.7	788	2	US-08-907-166-6	Sequence 6, Appl
796	6	0.7	799	4	US-09-180-439-6	Sequence 6, Appl
797	6	0.7	806	4	US-08-945-983-2	Sequence 2, Appl
798	6	0.7	809	4	US-09-105-537-24	Sequence 24, Appl
799	6	0.7	812	1	US-08-248-629A-1	Sequence 1, Appl
800	6	0.7	812	1	US-08-451-932-1	Sequence 1, Appl
801	6	0.7	812	1	US-08-452-960-1	Sequence 1, Appl
802	6	0.7	812	2	US-08-326-785-1	Sequence 1, Appl
803	6	0.7	812	2	US-08-612-788-1	Sequence 1, Appl
804	6	0.7	812	2	US-08-605-598B-1	Sequence 1, Appl
805	6	0.7	812	2	US-08-429-743-1	Sequence 1, Appl
806	6	0.7	812	2	US-08-866-735-1	Sequence 1, Appl
807	6	0.7	812	3	PCT-US95-05107-1	Sequence 1, Appl
808	6	0.7	819	4	US-09-651-656-15	Sequence 15, Appl
809	6	0.7	819	4	US-09-650-855-15	Sequence 15, Appl
810	6	0.7	821	1	US-07-935-311A-4	Sequence 4, Appl
811	6	0.7	821	1	US-08-368-079-4	Sequence 4, Appl
812	6	0.7	821	5	PCT-US93-07999-4	Sequence 4, Appl
813	6	0.7	830	4	US-09-562-737-36	Sequence 36, Appl
814	6	0.7	837	4	US-09-122-126B-2	Sequence 2, Appl
815	6	0.7	842	2	US-08-928-692-18	Sequence 18, Appl
816	6	0.7	842	3	US-09-339-972-18	Sequence 22, Appl
817	6	0.7	843	3	US-09-215-966-22	Sequence 22, Appl
818	6	0.7	845	1	US-08-416-950-11	Sequence 11, Appl
819	6	0.7	845	2	US-08-469-830-11	Sequence 11, Appl
820	6	0.7	852	1	US-08-190-802A-59	Sequence 59, Appl
821	6	0.7	852	1	US-08-190-802A-67	Sequence 67, Appl
822	6	0.7	852	4	US-08-477-346-59	Sequence 59, Appl
823	6	0.7	852	4	US-08-477-346-67	Sequence 67, Appl
824	6	0.7	852	4	US-08-473-089-59	Sequence 59, Appl
825	6	0.7	852	4	US-08-487-072A-67	Sequence 67, Appl
826	6	0.7	852	4	US-08-487-072A-67	Sequence 67, Appl
827	6	0.7	852	4	US-08-487-072A-67	Sequence 67, Appl
828	6	0.7	861	3	US-09-022-875-2	Sequence 2, Appl
829	6	0.7	863	3	US-08-666-271-2	Sequence 2, Appl
830	6	0.7	876	2	US-08-510-215A-2	Sequence 2, Appl

831	6	0.7	876	2	US-08-633-476-2	Sequence 2, Appll	904	6	0.7	1130	2	US-08-125-077-2	Sequence 2, Appll
832	6	0.7	876	2	US-08-436-664-20	Sequence 30, Appl	905	6	0.7	1130	6	5444158-2	Patent No. 5444158
833	6	0.7	876	2	US-08-436-664-32	Sequence 32, Appl	906	6	0.7	1130	6	US-09-238-303-9	Sequence 9, Appll
834	6	0.7	876	2	US-08-436-664-34	Sequence 34, Appl	907	6	0.7	1155	1	US-08-286-889-46	Sequence 46, Appl
835	6	0.7	876	3	US-09-135-642-20	Sequence 20, Appl	908	6	0.7	1155	1	US-08-485-618-46	Sequence 46, Appl
836	6	0.7	876	3	US-09-135-642-32	Sequence 32, Appl	909	6	0.7	1155	1	US-08-362-652-46	Sequence 46, Appl
837	6	0.7	876	3	US-09-135-642-34	Sequence 34, Appl	910	6	0.7	1155	2	US-08-605-672-46	Sequence 46, Appl
838	6	0.7	876	3	US-08-394-232A-20	Sequence 20, Appl	911	6	0.7	1155	2	US-08-482-293A-46	Sequence 46, Appl
839	6	0.7	876	3	US-08-394-232A-32	Sequence 32, Appl	912	6	0.7	1155	2	US-08-482-293A-46	Sequence 46, Appl
840	6	0.7	876	3	US-08-394-232A-34	Sequence 34, Appl	913	6	0.7	1155	2	US-09-193-043-46	Sequence 46, Appl
841	6	0.7	876	4	US-09-517-871-2	Sequence 2, Appll	914	6	0.7	1155	4	US-09-688-307A-46	Sequence 46, Appl
842	6	0.7	876	4	US-09-517-871-20	Sequence 20, Appl	915	6	0.7	1156	4	US-09-002-285-72	Sequence 72, Appl
843	6	0.7	876	5	PCT-US95-04080-20	Sequence 20, Appl	916	6	0.7	1157	2	US-08-532-547-5	Sequence 5, Appll
844	6	0.7	876	5	PCT-US95-04080-32	Sequence 32, Appl	917	6	0.7	1157	2	US-08-379-656B-5	Sequence 5, Appll
845	6	0.7	876	5	PCT-US95-04080-34	Sequence 34, Appl	918	6	0.7	1157	3	US-08-455-838-5	Sequence 5, Appll
846	6	0.7	877	1	US-08-208-036-14	Sequence 14, Appl	919	6	0.7	1157	4	US-09-019-809-5	Sequence 5, Appll
847	6	0.7	877	1	US-08-428-823-14	Sequence 14, Appl	920	6	0.7	1157	4	US-09-471-177-5	Sequence 5, Appll
848	6	0.7	906	2	US-08-609-230A-9	Sequence 9, Appll	921	6	0.7	1161	1	US-08-485-618-53	Sequence 53, Appl
849	6	0.7	907	3	US-08-990-140-4	Sequence 4, Appll	922	6	0.7	1161	1	US-08-362-652-53	Sequence 53, Appl
850	6	0.7	907	4	US-09-546-238-4	Sequence 4, Appll	923	6	0.7	1161	2	US-08-605-672-53	Sequence 53, Appl
851	6	0.7	913	1	US-08-487-890A-5	Sequence 5, Appll	924	6	0.7	1161	2	US-08-482-293A-53	Sequence 53, Appl
852	6	0.7	913	1	US-08-445-640-4	Sequence 4, Appll	925	6	0.7	1161	2	US-08-943-363-53	Sequence 53, Appl
853	6	0.7	913	2	US-08-478-435-5	Sequence 5, Appll	926	6	0.7	1161	4	US-09-193-043-53	Sequence 53, Appl
854	6	0.7	913	2	US-08-337-483-5	Sequence 5, Appll	927	6	0.7	1161	4	US-09-588-307A-53	Sequence 53, Appl
855	6	0.7	913	2	US-08-478-373-5	Sequence 5, Appll	928	6	0.7	1169	2	US-08-542-921-2	Sequence 2, Appll
856	6	0.7	913	3	US-08-170-558-4	Sequence 4, Appll	929	6	0.7	1169	2	US-08-880-685-2	Sequence 2, Appll
857	6	0.7	913	3	US-08-474-671-5	Sequence 5, Appll	930	6	0.7	1169	2	US-08-880-685-2	Sequence 2, Appll
858	6	0.7	913	3	US-08-483-577A-5	Sequence 5, Appll	931	6	0.7	1182	4	US-09-041-886-21	Sequence 21, Appl
859	6	0.7	913	3	US-08-447-314-4	Sequence 4, Appll	932	6	0.7	1185	4	US-09-134-001C-5276	Sequence 5276, Ap
860	6	0.7	913	3	US-08-445-461-4	Sequence 4, Appll	933	6	0.7	1207	1	US-07-951-715A-7	Sequence 7, Appll
861	6	0.7	913	4	US-08-897-438-5	Sequence 5, Appll	934	6	0.7	1207	3	US-08-459-448A-7	Sequence 7, Appll
862	6	0.7	913	4	US-08-637-654-5	Sequence 5, Appll	935	6	0.7	1207	3	US-08-459-595A-7	Sequence 7, Appll
863	6	0.7	913	4	US-08-649-518-5	Sequence 5, Appll	936	6	0.7	1207	3	US-08-459-504B-7	Sequence 7, Appll
864	6	0.7	917	4	US-09-049-698-41	Sequence 41, Appl	937	6	0.7	1207	3	US-08-459-444-7	Sequence 7, Appll
865	6	0.7	919	1	US-08-336-343A-2	Sequence 2, Appll	938	6	0.7	1207	3	US-09-053-549-8	Sequence 8, Appll
866	6	0.7	936	5	PCT-US93-05944-2	Sequence 2, Appll	939	6	0.7	1207	4	US-09-547-422-7	Sequence 7, Appll
867	6	0.7	940	4	US-08-810-712-7	Sequence 7, Appll	940	6	0.7	1213	1	US-08-188-582-20	Sequence 20, Appl
868	6	0.7	947	4	US-09-228-986-73	Sequence 73, Appl	941	6	0.7	1213	1	US-08-646-715-20	Sequence 20, Appl
869	6	0.7	954	3	US-09-057-969-2	Sequence 2, Appll	942	6	0.7	1227	1	US-08-448-170-8	Sequence 8, Appll
870	6	0.7	956	2	US-08-897-443-3	Sequence 3, Appll	943	6	0.7	1227	3	US-09-053-549-2	Sequence 2, Appll
871	6	0.7	956	4	US-09-134-001C-4452	Sequence 4452, Ap	944	6	0.7	1227	4	US-08-961-803-9	Sequence 9, Appll
872	6	0.7	968	4	US-09-180-439-3	Sequence 3, Appll	945	6	0.7	1262	4	US-09-357-251-33	Sequence 33, Appl
873	6	0.7	968	4	US-09-180-439-4	Sequence 4, Appll	946	6	0.7	1266	1	US-08-468-557-4	Sequence 4, Appll
874	6	0.7	968	4	US-09-228-986-76	Sequence 76, Appl	947	6	0.7	1266	1	US-09-357-251-32	Sequence 32, Appl
875	6	0.7	979	4	US-09-323-872A-55	Sequence 55, Appl	948	6	0.7	1306	4	US-08-999-774A-13	Sequence 13, Appl
876	6	0.7	983	4	US-09-134-001C-3814	Sequence 3814, Ap	949	6	0.7	1333	4	US-09-347-878-20	Sequence 20, Appl
877	6	0.7	984	1	US-08-257-073-3	Sequence 3, Appll	950	6	0.7	1365	4	US-09-376-330-18	Sequence 18, Appl
878	6	0.7	984	2	US-08-184-009-120	Sequence 120, App	951	6	0.7	1365	6	5194600-4	Patent No. 5194600
879	6	0.7	984	2	US-08-458-356-120	Sequence 120, App	952	6	0.7	1404	2	US-08-400-159-2	Sequence 2, Appll
880	6	0.7	984	4	US-08-460-736-120	Sequence 120, App	953	6	0.7	1404	3	US-08-611-729A-2	Sequence 2, Appll
881	6	0.7	988	2	US-08-286-819A-19	Sequence 19, App	954	6	0.7	1422	4	US-08-469-260A-85	Sequence 85, Appl
882	6	0.7	988	3	US-08-980-357-19	Sequence 19, Appl	955	6	0.7	1430	3	US-09-008-172-2	Sequence 2, Appll
883	6	0.7	994	3	US-08-699-103B-3	Sequence 3, Appll	956	6	0.7	1430	4	US-09-210-361-6	Sequence 6, Appll
884	6	0.7	994	4	US-09-229-059-3	Sequence 3, Appll	957	6	0.7	1430	4	US-09-740-274-6	Sequence 6, Appll
885	6	0.7	1004	4	US-09-268-347-30	Sequence 30, Appl	958	6	0.7	1528	1	US-08-463-092B-6	Sequence 6, Appll
886	6	0.7	1009	2	US-08-357-642A-1	Sequence 1, Appll	959	6	0.7	1528	2	US-08-462-109A-6	Sequence 6, Appll
887	6	0.7	1009	2	US-08-460-626-1	Sequence 1, Appll	960	6	0.7	1528	3	US-08-460-907B-6	Sequence 6, Appll
888	6	0.7	1016	4	US-09-180-439-8	Sequence 8, Appll	961	6	0.7	1528	3	US-08-463-179A-6	Sequence 6, Appll
889	6	0.7	1024	4	US-09-562-737-48	Sequence 48, Appl	962	6	0.7	1528	3	US-08-463-179A-6	Sequence 6, Appll
890	6	0.7	1031	4	US-08-811-682-15	Sequence 15, Appl	963	6	0.7	1528	3	US-08-461-384B-6	Sequence 6, Appll
891	6	0.7	1044	2	US-08-777-405A-2	Sequence 2, Appll	964	6	0.7	1531	1	US-08-141-893-3	Sequence 4, Appll
892	6	0.7	1044	2	US-08-977-871A-2	Sequence 2, Appll	965	6	0.7	1531	1	US-08-463-092B-2	Sequence 2, Appll
893	6	0.7	1044	2	US-09-225-951-2	Sequence 2, Appll	966	6	0.7	1531	1	US-08-463-092B-2	Sequence 2, Appll
894	6	0.7	1048	4	US-09-171-699-10	Sequence 10, Appl	967	6	0.7	1531	2	US-08-462-109A-4	Sequence 4, Appll
895	6	0.7	1048	4	US-08-887-534A-85	Sequence 85, Appl	968	6	0.7	1531	2	US-08-460-907B-2	Sequence 2, Appll
896	6	0.7	1073	4	US-09-206-942-49	Sequence 49, Appl	969	6	0.7	1531	2	US-08-460-907B-2	Sequence 2, Appll
897	6	0.7	1079	4	US-09-206-942-47	Sequence 47, Appl	970	6	0.7	1531	3	US-08-463-179A-2	Sequence 2, Appll
898	6	0.7	1104	4	US-09-268-347-28	Sequence 28, Appl	971	6	0.7	1531	3	US-08-463-179A-2	Sequence 2, Appll
899	6	0.7	1104	4	US-09-268-347-34	Sequence 34, Appl	972	6	0.7	1531	3	US-08-461-384B-2	Sequence 2, Appll
900	6	0.7	1112	4	US-09-353-585-2	Sequence 2, Appll	973	6	0.7	1531	3	US-08-461-384B-2	Sequence 2, Appll
901	6	0.7	1112	4	US-09-353-585-3	Sequence 3, Appll	974	6	0.7	1531	3	US-08-407-207A-2	Sequence 2, Appll
902	6	0.7	1118	4	US-09-379-523-3	Sequence 3, Appll	975	6	0.7	1566	6	US-08-687-956A-23	Sequence 23, Appl
903	6	0.7	1130	2	US-08-460-309-2	Sequence 2, Appll	976	6	0.7	1676	4	US-08-487-283A-2	Sequence 2, Appll


```

: APPLICANT: Wen, Zilong
: TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
: TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Klauber & Jackson
: STREET: 411 Hackensack Avenue
: CITY: Hackensack
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/820,754
: FILING DATE: 19-MAR-1997
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/212,185
: FILING DATE: 11-MAR-1994
: APPLICATION NUMBER: US 07/980,498
: FILING DATE: 23-NOV-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/854,296
: FILING DATE: 19-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO US93/02569
: FILING DATE: 19-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/126,588
: FILING DATE: 24-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Jackson Esq., David A.
: REGISTRATION NUMBER: 26,742
: REFERENCE/DOCKET NUMBER: 600-1-073 CIP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201 487-5800
: TELEFAX: 201 343-1684
: TELEX: 133521
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 851 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-820-754-2
:
: Query Match 1.0%; Score 9; DB 2; Length 851;
: Best Local Similarity 100.0%; Pred. No. 3;
: Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: Oy 457 ESTLEPVIE 465
: Db 756 ESTLEPVIE 764
:
: RESULT 4
: US-08-956-652-2
: Sequence 2, Application US/08956652
: Patent No. 6013475
: GENERAL INFORMATION:
: APPLICANT: Darnell Jr., James E.
: APPLICANT: Schindler, Christian W.
: APPLICANT: Fu, Xian-Yuan
: APPLICANT: Wen, Zilong
: APPLICANT: Zhong, Zhong
: TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
: TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
: NUMBER OF SEQUENCES: 25

```

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Klauber & Jackson
: STREET: 411 Hackensack Avenue
: CITY: Hackensack
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/956,652
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/212,185
: FILING DATE: 11-MAR-1994
: APPLICATION NUMBER: US 07/980,498
: FILING DATE: 23-NOV-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/854,296
: FILING DATE: 19-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO US93/02569
: FILING DATE: 19-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/126,588
: FILING DATE: 24-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Jackson Esq., David A.
: REGISTRATION NUMBER: 26,742
: REFERENCE/DOCKET NUMBER: 600-1-073 CIP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201 487-5800
: TELEFAX: 201 343-1684
: TELEX: 133521
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 851 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-956-652-2
:
: Query Match 1.0%; Score 9; DB 3; Length 851;
: Best Local Similarity 100.0%; Pred. No. 3;
: Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: Oy 457 ESTLEPVIE 465
: Db 756 ESTLEPVIE 764
:
: RESULT 5
: US-08-956-869-2
: Sequence 2, Application US/08956689
: Patent No. 6030808
: GENERAL INFORMATION:
: APPLICANT: Darnell Jr., James E.
: APPLICANT: Schindler, Christian W.
: APPLICANT: Fu, Xian-Yuan
: APPLICANT: Wen, Zilong
: APPLICANT: Zhong, Zhong
: TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
: TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Klauber & Jackson
: STREET: 411 Hackensack Avenue
: CITY: Hackensack
: STATE: New Jersey

```

COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,869
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/212,185
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 851 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-956-869-2

Query Match 1.0%; Score 9; DB 3; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465
|||||||
Db 756 ESTLEPVIE 764

RESULT 6
US-09-012-710-2
Sequence 2, Application US/09012710
Patent No. 6087478
GENERAL INFORMATION:
APPLICANT: Vinkemeler, Uwe
APPLICANT: Morelli, Ismail
APPLICANT: Darnell, Jr., James E.
APPLICANT: Kuriyan, John
TITLE OF INVENTION: A CRYSTAL OF THE N-TERMINAL DOMAIN OF A
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/012,710
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-194
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 851 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-012-710-2

Query Match 1.0%; Score 9; DB 3; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465
|||||||
Db 756 ESTLEPVIE 764

RESULT 7
US-08-948-547-2
Sequence 2, Application US/08948547
Patent No. 6124118
GENERAL INFORMATION:
APPLICANT: Darnell Jr., James E.
APPLICANT: Schindler, Christian W.
APPLICANT: Fu, Xian-Yuan
APPLICANT: Wen, Zilong
APPLICANT: Zhong, Zhong
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESS: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,547
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/212,185
FILING DATE: 11-MAR-1994
APPLICATION NUMBER: US 07/980,498
FILING DATE: 23-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588
FILING DATE: 24-SEP-1993


```
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 851 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-948-547-2

Query Match      1.0%; Score 9; DB 3; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465
   |||||||
Db 756 ESTLEPVIE 764

RESULT 8
US-09-087-465-4
; Sequence 4, Application US/09087465A
; Patent No. 6160092
; GENERAL INFORMATION:
; APPLICANT: Vinkemeier, Uwe
; APPLICANT: Chen, Xiaomin
; APPLICANT: Darnell Jr., James E
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF THE CORE PORTION OF A STAT AND METHODS OF
; FILE REFERENCE: 600-1-229
; CURRENT APPLICATION NUMBER: US/09/087,465A
; CURRENT FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 851
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-087-465-4

Query Match      1.0%; Score 9; DB 4; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465
   |||||||
Db 756 ESTLEPVIE 764

RESULT 9
US-09-364-970-2
; Sequence 2, Application US/09364970
; Patent No. 6235873
; GENERAL INFORMATION:
; APPLICANT: Bromberg, Jacqueline
; TITLE OF INVENTION: CONSTITUTIVELY ACTIVE TRANSCRIPTION FACTORS AND THEIR
; TITLE OF INVENTION: USES FOR IDENTIFYING MODULATORS OF ACTIVITY INCLUDING
; FILE REFERENCE: 600-1-252
; CURRENT APPLICATION NUMBER: US/09/364,970
; CURRENT FILING DATE: 1999-07-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 851
; TYPE: PRT
```

```
ORGANISM: Homo sapiens
US-09-364-970-2

Query Match      1.0%; Score 9; DB 4; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465
   |||||||
Db 756 ESTLEPVIE 764

RESULT 10
US-09-556-273-2
; Sequence 2, Application US/09556273
; Patent No. 6312887
; GENERAL INFORMATION:
; APPLICANT: Vinkemeier, Uwe
; APPLICANT: Marell, Ismail
; APPLICANT: Darnell, Jr., James E.
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF THE N-TERMINAL DOMAIN OF A
; TITLE OF INVENTION: STAT PROTEIN AND METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/556,273
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/012,710
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-194
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 851 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-556-273-2

Query Match      1.0%; Score 9; DB 4; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465
   |||||||
Db 756 ESTLEPVIE 764

RESULT 11
US-08-956-653A-2
; Sequence 2, Application US/08956653A
```

Patent No. 6338949
GENERAL INFORMATION:
APPLICANT: Darnell Jr., James E.
APPLICANT: Schindler, Christian W.
APPLICANT: Fu, Xian-Yuan
APPLICANT: Wen, Zilong
APPLICANT: Zhong, Zhong
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,653A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,185
FILING DATE: 11-MAR-1994
APPLICATION NUMBER: US 07/980,498
FILING DATE: 23-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-195
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 851 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-956-653A-2

Query Match 1.0%; Score 9; DB 4; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ESTLEPVIE 465
DB 756 ESTLEPVIE 764

RESULT 12
PCT-US95-17025-2
Sequence 2, Application PC/TUS9517025
GENERAL INFORMATION:
APPLICANT: James E. Darnell, Jr.
APPLICANT: Zilong Wen
APPLICANT: Curt M. Horvath
APPLICANT: Zhong Zhong

TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL
TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/17025
FILING DATE: 28-DEC-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/369,796
FILING DATE: 06-JAN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-116
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 851 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-17025-2

Query Match 1.0%; Score 9; DB 5; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ESTLEPVIE 465
DB 756 ESTLEPVIE 764

RESULT 13
US-08-276-099A-13
Sequence 13, Application US/08276099A
Patent No. 5591825
GENERAL INFORMATION:
APPLICANT: McKnight, Steven L
APPLICANT: Hou, Jinzhao
TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
TITLE OF INVENTION: BINDING ASSAYS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HORBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,099A
FILING DATE: 15-JUL-1994

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59451-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 852 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-276-099A-13

Query Match 1.0%; Score 9; DB 1; Length 852;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ESTLEPVIE 465
|||||
DB 756 ESTLEPVIE 764

RESULT 14
US-08-781-890-13
Sequence 13, Application US/08781890
Patent No. 5710266
GENERAL INFORMATION:
APPLICANT: McKnight, Steven L
APPLICANT: Hou, Jinzhao
TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
TITLE OF INVENTION: BINDING ASSAYS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,890
FILING DATE: 05-JAN-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/276,099
FILING DATE: 15-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard Aron
REGISTRATION NUMBER: 36,627
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 852 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-781-890-13

Query Match 1.0%; Score 9; DB 1; Length 852;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ESTLEPVIE 465
|||||
DB 756 ESTLEPVIE 764

RESULT 15
US-09-242-632A-2
Sequence 2, Application US/09242632A
Patent No. 6312696
GENERAL INFORMATION:
APPLICANT: Nippon Zeon
TITLE OF INVENTION: Antigen protein derived from infectious
TITLE OF INVENTION: Laryngotracheitis virus
FILE REFERENCE: E868-PCF
CURRENT APPLICATION NUMBER: US/09/242,632A
CURRENT FILING DATE: 1999-02-19
PRIOR APPLICATION NUMBER: JP 8-238580
PRIOR FILING DATE: 1996-08-21
NUMBER OF SEQ ID NOS: 28
SEQ ID NO 2
LENGTH: 582
TYPE: PRT
ORGANISM: Infectious laryngotracheitis virus
FEATURE:
OTHER INFORMATION: Amino acid sequence of 132h polypeptide of
OTHER INFORMATION: Infectious laryngotracheitis virus
US-09-242-632A-2

Query Match 0.9%; Score 8; DB 4; Length 582;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 213 KAALEDIT 220
|||||
DB 277 KAALEDIT 284

RESULT 16
US-07-960-510-4
Sequence 4, Application US/07960510
Patent No. 5705362
GENERAL INFORMATION:
APPLICANT: BONEKAMP, ALFONSUSJOHANNES
APPLICANT: VAN TILBORG E., MARCELIS W.
TITLE OF INVENTION: MODIFIED SIGNAL SEQUENCES
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave., Suite 5500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/960,510
FILING DATE: 19921013
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 24615-20034.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763

TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-960-510-4

Query Match
Best Local Similarity 0.8%; Score 7; DB 1; Length 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 KRLVRL 172
DB 5 KRLVRL 11

RESULT 17
US-07-960-510-5
Sequence 5, Application US/07960510
Patent No. 5705362

GENERAL INFORMATION:
APPLICANT: BONEKAMP, ALFONSUSJOHANNES
APPLICANT: VAN TILBORG E., MARCELLIS W.
TITLE OF INVENTION: MODIFIED SIGNAL SEQUENCES
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave., Suite 5500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006-1888

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/960,510
FILING DATE: 19921013
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 24615-20034.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-960-510-5

Query Match
Best Local Similarity 0.8%; Score 7; DB 1; Length 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 KRLVRL 172
DB 5 KRLVRL 11

RESULT 18
US-07-960-510-6
Sequence 6, Application US/07960510
Patent No. 5705362
GENERAL INFORMATION:

APPLICANT: BONEKAMP, ALFONSUSJOHANNES
APPLICANT: VAN TILBORG E., MARCELLIS W.
TITLE OF INVENTION: MODIFIED SIGNAL SEQUENCES
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave., Suite 5500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006-1888

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/960,510
FILING DATE: 19921013
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 24615-20034.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-960-510-6

Query Match
Best Local Similarity 0.8%; Score 7; DB 1; Length 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 KRLVRL 172
DB 5 KRLVRL 11

RESULT 19
US-07-960-510-7
Sequence 7, Application US/07960510
Patent No. 5705362
GENERAL INFORMATION:

APPLICANT: BONEKAMP, ALFONSUSJOHANNES
APPLICANT: VAN TILBORG E., MARCELLIS W.
TITLE OF INVENTION: MODIFIED SIGNAL SEQUENCES
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave., Suite 5500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006-1888

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/960,510
FILING DATE: 19921013
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959

```
REFERENCE/DOCKET NUMBER: 24615-20034.00
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-960-510-7

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 166 KRLYARL 172
Db 5 KRLYARL 11

RESULT 20
5187089-8
Patent No. 5187089
APPLICANT: SCOTT, RANDY W.; GOLINI, FRED; MCGROGAN, MICHAEL
TITLE OF INVENTION: PROTEASE NEXIN-I VARIANTS WHICH INHIBIT ELASTASE
NUMBER OF SEQUENCES: 11
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/542,484
FILING DATE: 21-JUN-1990
SEQ ID NO: 8
LENGTH: 34
5187089-8

Query Match
Best Local Similarity 100.0%; Score 7; DB 6; Length 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 117 PLSTLEEL 123
Db 5 PLSTLEEL 11

RESULT 21
5457090-5
Patent No. 5457090
APPLICANT: SCOTT, RANDY W.; GOLINI, FRED; MCGROGAN, MICHAEL
TITLE OF INVENTION: PROTEASE NEXIN-I VARIANTS
NUMBER OF SEQUENCES: 7
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/924,294
FILING DATE: 03-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 342,484
FILING DATE: 21-JUN-1990
SEQ ID NO: 5
LENGTH: 34
5457090-5

Query Match
Best Local Similarity 100.0%; Score 7; DB 6; Length 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 117 PLSTLEEL 123
Db 5 PLSTLEEL 11

RESULT 22
5495001-16
Patent No. 5495001
```

```
APPLICANT: MCGROGAN, MICHAEL P.; SCOTT, RANDY W.; BAKER,
JOFFRE B.; SIMONSEN, CHRISTIAN C.
TITLE OF INVENTION: RECOMBINANT PURIFIED PROTEASE NEXIN
NUMBER OF SEQUENCES: 16
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/288,596
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 111,111
FILING DATE: 24-AUG-1993
APPLICATION NUMBER: 25,450
FILING DATE: 13-MAR-1987
APPLICATION NUMBER: 871,501
FILING DATE: 06-JUN-1986
APPLICATION NUMBER: 870,232
FILING DATE: 03-JUN-1986
SEQ ID NO: 16
LENGTH: 34
5495001-16

Query Match
Best Local Similarity 100.0%; Score 7; DB 6; Length 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 117 PLSTLEEL 123
Db 5 PLSTLEEL 11

RESULT 23
5171673-7
Patent No. 5171673
APPLICANT: SLOMA, ALAN; HANNETT, NANCY M.; STEPHENS, M.A.
RUDOLPH, CATHY F.; RUFO JR., GERALD A.; PERO, JANICE
TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS DNA USING THE BACILLUS COAGULANS AMYLASE GENE
NUMBER OF SEQUENCES: 10
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/219,599
FILING DATE: 18-JUL-1988
SEQ ID NO: 7
LENGTH: 35
5171673-7

Query Match
Best Local Similarity 100.0%; Score 7; DB 6; Length 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 166 KRLYARL 172
Db 5 KRLYARL 11

RESULT 24
US-09-149-476-601
Sequence 601, Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: EP002P1
CURRENT APPLICATION NUMBER: US/09/149,476
CURRENT FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
```

1	EARLIER APPLICATION NUMBER: 60/040,334
2	EARLIER FILING DATE: 1997-03-07
3	EARLIER APPLICATION NUMBER: 60/040,336
4	EARLIER FILING DATE: 1997-03-07
5	EARLIER APPLICATION NUMBER: 60/040,163
6	EARLIER FILING DATE: 1997-03-07
7	EARLIER APPLICATION NUMBER: 60/047,600
8	EARLIER FILING DATE: 1997-05-23
9	EARLIER APPLICATION NUMBER: 60/047,615
10	EARLIER FILING DATE: 1997-05-23
11	EARLIER APPLICATION NUMBER: 60/047,597
12	EARLIER FILING DATE: 1997-05-23
13	EARLIER APPLICATION NUMBER: 60/047,502
14	EARLIER FILING DATE: 1997-05-23
15	EARLIER APPLICATION NUMBER: 60/047,633
16	EARLIER FILING DATE: 1997-05-23
17	EARLIER APPLICATION NUMBER: 60/047,583
18	EARLIER FILING DATE: 1997-05-23
19	EARLIER APPLICATION NUMBER: 60/047,617
20	EARLIER FILING DATE: 1997-05-23
21	EARLIER APPLICATION NUMBER: 60/047,618
22	EARLIER FILING DATE: 1997-05-23
23	EARLIER APPLICATION NUMBER: 60/047,503
24	EARLIER FILING DATE: 1997-05-23
25	EARLIER APPLICATION NUMBER: 60/047,592
26	EARLIER FILING DATE: 1997-05-23
27	EARLIER APPLICATION NUMBER: 60/047,581
28	EARLIER FILING DATE: 1997-05-23
29	EARLIER APPLICATION NUMBER: 60/047,584
30	EARLIER FILING DATE: 1997-05-23
31	EARLIER APPLICATION NUMBER: 60/047,500
32	EARLIER FILING DATE: 1997-05-23
33	EARLIER APPLICATION NUMBER: 60/047,587
34	EARLIER FILING DATE: 1997-05-23
35	EARLIER APPLICATION NUMBER: 60/047,492
36	EARLIER FILING DATE: 1997-05-23
37	EARLIER APPLICATION NUMBER: 60/047,598
38	EARLIER FILING DATE: 1997-05-23
39	EARLIER APPLICATION NUMBER: 60/047,613
40	EARLIER FILING DATE: 1997-05-23
41	EARLIER APPLICATION NUMBER: 60/047,582
42	EARLIER FILING DATE: 1997-05-23
43	EARLIER APPLICATION NUMBER: 60/047,596
44	EARLIER FILING DATE: 1997-05-23
45	EARLIER APPLICATION NUMBER: 60/047,612
46	EARLIER FILING DATE: 1997-05-23
47	EARLIER APPLICATION NUMBER: 60/047,632
48	EARLIER FILING DATE: 1997-05-23
49	EARLIER APPLICATION NUMBER: 60/047,601
50	EARLIER FILING DATE: 1997-05-23
51	EARLIER APPLICATION NUMBER: 60/043,580
52	EARLIER FILING DATE: 1997-04-11
53	EARLIER APPLICATION NUMBER: 60/043,568
54	EARLIER FILING DATE: 1997-04-11
55	EARLIER APPLICATION NUMBER: 60/043,314
56	EARLIER FILING DATE: 1997-04-11
57	EARLIER APPLICATION NUMBER: 60/043,569
58	EARLIER FILING DATE: 1997-04-11
59	EARLIER APPLICATION NUMBER: 60/043,311
60	EARLIER FILING DATE: 1997-04-11
61	EARLIER APPLICATION NUMBER: 60/043,671
62	EARLIER FILING DATE: 1997-04-11
63	EARLIER APPLICATION NUMBER: 60/043,674
64	EARLIER FILING DATE: 1997-04-11
65	EARLIER APPLICATION NUMBER: 60/043,669
66	EARLIER FILING DATE: 1997-04-11
67	EARLIER APPLICATION NUMBER: 60/043,312
68	EARLIER FILING DATE: 1997-04-11
69	EARLIER APPLICATION NUMBER: 60/043,313
70	EARLIER FILING DATE: 1997-04-11
71	EARLIER APPLICATION NUMBER: 60/043,672
72	EARLIER FILING DATE: 1997-04-11
73	EARLIER APPLICATION NUMBER: 60/043,315
74	EARLIER FILING DATE: 1997-04-11
75	EARLIER APPLICATION NUMBER: 60/048,974
76	EARLIER FILING DATE: 1997-06-06
77	EARLIER APPLICATION NUMBER: 60/056,886
78	EARLIER FILING DATE: 1997-08-22
79	EARLIER APPLICATION NUMBER: 60/056,877
80	EARLIER FILING DATE: 1997-08-22
81	EARLIER APPLICATION NUMBER: 60/056,889
82	EARLIER FILING DATE: 1997-08-22
83	EARLIER APPLICATION NUMBER: 60/056,893
84	EARLIER FILING DATE: 1997-08-22
85	EARLIER APPLICATION NUMBER: 60/056,630
86	EARLIER FILING DATE: 1997-08-22
87	EARLIER APPLICATION NUMBER: 60/056,878
88	EARLIER FILING DATE: 1997-08-22
89	EARLIER APPLICATION NUMBER: 60/056,662
90	EARLIER FILING DATE: 1997-08-22
91	EARLIER APPLICATION NUMBER: 60/056,872
92	EARLIER FILING DATE: 1997-08-22
93	EARLIER APPLICATION NUMBER: 60/056,882
94	EARLIER FILING DATE: 1997-08-22
95	EARLIER APPLICATION NUMBER: 60/056,637
96	EARLIER FILING DATE: 1997-08-22
97	EARLIER APPLICATION NUMBER: 60/056,903
98	EARLIER FILING DATE: 1997-08-22
99	EARLIER APPLICATION NUMBER: 60/056,894
100	EARLIER FILING DATE: 1997-08-22
101	EARLIER APPLICATION NUMBER: 60/056,911
102	EARLIER FILING DATE: 1997-08-22
103	EARLIER APPLICATION NUMBER: 60/056,636
104	EARLIER FILING DATE: 1997-08-22
105	EARLIER APPLICATION NUMBER: 60/056,874
106	EARLIER FILING DATE: 1997-08-22
107	EARLIER APPLICATION NUMBER: 60/056,910
108	EARLIER FILING DATE: 1997-08-22
109	EARLIER APPLICATION NUMBER: 60/056,864
110	EARLIER FILING DATE: 1997-08-22
111	EARLIER APPLICATION NUMBER: 60/056,631
112	EARLIER FILING DATE: 1997-08-22
113	EARLIER APPLICATION NUMBER: 60/056,845
114	EARLIER FILING DATE: 1997-08-22
115	EARLIER APPLICATION NUMBER: 60/056,892
116	EARLIER FILING DATE: 1997-08-22
117	EARLIER APPLICATION NUMBER: 60/057,761
118	EARLIER FILING DATE: 1997-08-22
119	

```

; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13
; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

Query Match      0.8%; Score 7; DB 4; Length 69;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 ELLEOLL 394
DB 51 ELLEOLL 57

RESULT 25
US-09-149-476-465
; Sequence 465, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002pi
; CURRENT FILING DATE: US/09/149,476
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
```

```

; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
```

```

EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,879
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,880
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,894
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,911
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,636
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,874
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,910
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,864
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,892
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,761
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/047,595
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,599
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,588
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,585
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,586
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,590
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,594
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,589
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,593
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,614
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,578
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,670
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056,632

```

```

EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,664
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049,610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061,060
EARLIER FILING DATE: 1997-10-02

```

```

Query Match      0.8%; Score 7; DB 4; Length 70;
Best local similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 388 ELLEQLL 394
Db 51 ELLEQLL 57

```

```

RESULT 26
US-08-438-753B-24
; Sequence 24, Application US/08438753B
; Patent No. 5705363
; GENERAL INFORMATION:
; APPLICANT: Imakawa, Kazuhito
; TITLE OF INVENTION: Interferon Tau Compositions and
; TITLE OF INVENTION: Methods of Use
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/438,753B
; FILING DATE: 10-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/139,891
; FILING DATE: 19-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/847,741
; FILING DATE: 09-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/318,050
; FILING DATE: 02-MAR-1989

```



```

983 6 0.7 517 1 MBEA_ECOLI      P13658 escherichia
984 6 0.7 517 1 Y652_HUMAN      O75143 homo sapien
985 6 0.7 518 1 ASB3_HUMAN      O95757 homo sapien
986 6 0.7 519 1 CP5J_CANNA      Q12588 candida mai
987 6 0.7 519 1 DHA5_YEAST      P40047 saccharomyc
988 6 0.7 520 1 CFP3_HUMAN      Q08477 homo sapien
989 6 0.7 520 1 C2CB_ALCEU      P13510 alcaigenes
990 6 0.7 520 1 C2CB_ALCEU      P94176 alcaigenes
991 6 0.7 521 1 IMAL_XENLA      P52170 xenopus lae
992 6 0.7 522 1 RECN_HELPIJ      O92180 helicobacte
993 6 0.7 523 1 CYP4_CAEBL      P52012 caenorhabdi
994 6 0.7 523 1 U121_PPNV3      Q00703 pseudorabie
995 6 0.7 524 1 G6PD_SYNP3      P29686 synechococc
996 6 0.7 524 1 NIKA_ECOLI      P13590 escherichia
997 6 0.7 524 1 P60_LISTIV      O01837 listeria iv
998 6 0.7 524 1 RECN_HELPIJ      O25943 helicobacte
999 6 0.7 525 1 ASB3_MOUSE      O94472 mus musculu
1000 6 0.7 526 1 BUTY_HUMAN      Q13410 homo sapien

```

ALIGNMENTS

RESULT 1

```

YTFM_HAEIN STANDARD: PRT: 578 AA.
ID YTFM_HAEIN
AC P44038;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein HI0698 precursor.
GN HI0698.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
NC NCBI_TaxID=727;
RX SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RA MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Scott J.D., Shirley R., Liu L.-I., Gilek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ullrich J., Fritchman J.L., Fritchman J.L., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fritchman J.L., Saudek D.M., Brandon R.C.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae."
RL Electrophoresis 21:411-429(2000).
CC -1- SIMILARITY: STRONG, TO E.COLI YTFM.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32752; AAC2357.1;
CC TIGR: HI0698;
CC InterPro: IPR000184; Bac_surfAg-D15.
CC Pfam: PF01103; Bac_surface_Ag; 1.

```

```

KW Signal: Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 578 PROTEIN HI0698.
SQ SEQUENCE 578 AA; 65812 MW; 34F9AC189C505876 CRC64;
Query Match 1.2%; Score 11; DB 1; Length 578;
Best Local Similarity 100.0%; Pred. No. 0.0072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 882 GVGVRNAPVPG 892
Db 539 GVGVRNAPVPG 549

```

RESULT 2

```

STA2_HUMAN STANDARD: PRT: 851 AA.
ID STA2_HUMAN
AC P52630; O16430; O16431;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Signal transducer and activator of transcription 2 (p113).
GN STAT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NC NCBI_TaxID=9606;
RX SEQUENCE FROM N.A.
RN [1]
RA MEDLINE=95192056; PubMed=7885841;
RA Yan R., Qureshi S., Zhong Z., Wen Z., Darnell J.E. Jr.;
RT "The genomic structure of the STAT genes: multiple exons in
RT coincident sites in Stat1 and Stat2."
RL Nucleic Acids Res. 23:459-463(1995).
RN [2]
RX SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RA MEDLINE=92366558; PubMed=1502204;
RA Fu X.-Y., Schindler C., Improbata T., Aebersold R., Darnell J.E. Jr.;
RT "The proteins of ISGF-3, the interferon alpha-induced transcriptional
RT activator, define a gene family involved in signal transduction."
RL Proc. Natl. Acad. Sci. U.S.A. 89:7840-7843(1992).
RN [3]
RX ALTERNATIVE SPLICING.
RA MEDLINE=96176320; PubMed=8601453;
RA Sugiyama T., Nishio Y., Kishimoto T., Akira S.;
RT "Identification of alternative splicing form of Stat2."
RL FEBS Lett. 381:191-194(1996).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=97172544; PubMed=9020188;
RA Bluyssen H.A., Levy D.E.;
RT "Stat2 is a transcriptional activator that requires sequence-specific
RT contacts provided by stat1 and p48 for stable interaction with DNA."
RL J Biol. Chem. 272:4600-4605(1997).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO THE IFN-STIMULATED
CC RESPONSE ELEMENT (ISRE) AND TO THE GAS ELEMENT. THIS MULTIPROTEIN
CC TRANSCRIPTION FACTOR IS TERMED ISRG3.
CC -1- SUBUNIT: IN RESPONSE TO IFN ALPHA/BETA, THREE SUBUNITS (STAT1-
CC ALPHA, STAT1-BETA, STAT2) OF ISRG3, BECOME PHOSPHORYLATED ON
CC TYROSINE, MIGRATE INTO THE NUCLEUS, AND ASSEMBLE INTO A COMPLEX
CC TOGETHER WITH ISRG3 GAMMA (P48), A DNA-BINDING PROTEIN THAT
CC SPECIFICALLY BINDS TO THE IFN-STIMULATED RESPONSE ELEMENT.
CC -1- SUBCELLULAR LOCATION: Nuclear; translocated into the nucleus in
CC response to phosphorylation.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PTM: TYROSINE PHOSPHORYLATED IN RESPONSE TO IFN-ALPHA.
CC -1- SIMILARITY: BELONGS TO THE STAT FAMILY OF TRANSCRIPTION FACTORS.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

837	6	0.7	439	1	YSCN_YERPS	P40291 yersinia ps	910	6	0.7	478	1	Y143_MYCLE	032912 mycobacteri
838	6	0.7	440	1	EX7L_RALSO	O8xwe7 ralstonia s	911	6	0.7	479	1	CATL_BOTCI	P55304 botrytis ci
839	6	0.7	440	1	LCAT_HUMAN	P04180 homo sapien	912	6	0.7	479	1	PRL2_ARATH	Q39190 arabidopsis
840	6	0.7	440	1	LCAT_PAPAN	Q08738 papio anuul	913	6	0.7	480	1	CFIA_HUMAN	O15519 h casp8 and
841	6	0.7	440	1	SOX4_MOUSE	Q06830 mus musculu	914	6	0.7	480	1	ORP2_HUMAN	O9h1p3 homo sapien
842	6	0.7	440	1	YHR1_YEAST	P38820 saccharomyc	915	6	0.7	480	1	Y486_MYCTU	O11152 mycobacteri
843	6	0.7	441	1	DHOM_MYCTU	Q10601 mycobacteri	916	6	0.7	481	1	GLGA_RHILLO	Q98562 rhizobium 1
844	6	0.7	441	1	FLUD_AOUAE	O67805 aquifex aeo	917	6	0.7	481	1	PRTB_ERWCH	P16316 erwina chr
845	6	0.7	442	1	CITN_LACLA	P21608 lactococcu	918	6	0.7	481	1	PTMB_CLOAB	O65969 clostridium
846	6	0.7	442	1	TBB_STYLE	P11857 stylonchias	919	6	0.7	481	1	XYLB_SRRRU	P27156 streptomyce
847	6	0.7	444	1	KRAQ_DICDI	P54644 dictyostell	920	6	0.7	482	1	Y269_TREPA	O83293 treponema p
848	6	0.7	445	1	GUDH_ECOLI	P76637 escherichia	921	6	0.7	483	1	YJQ2_YEAST	P46997 saccharomyc
849	6	0.7	446	1	NU4M_CERCA	Q34048 ceratilis c	922	6	0.7	483	1	GLGS_ORYSA	P15280 oryza sativ
850	6	0.7	446	1	NU4M_DROME	P18931 mus musculu	923	6	0.7	483	1	LYXK_PASMU	P57928 pasteurella
851	6	0.7	446	1	NU4M_DROVA	P07707 drosophila	924	6	0.7	483	1	MURE_CHLPN	Q928c5 chlamydia p
852	6	0.7	446	1	GN7L_RAT	O09325 rattus norv	925	6	0.7	483	1	MURE_CHLTR	O84271 chlamydia t
853	6	0.7	448	1	FBL5_HUMAN	Q9ubx5 homo sapien	926	6	0.7	485	1	GATP_BACST	Q931e2 bacillus st
854	6	0.7	448	1	FBL5_MOUSE	Q9wvh8 mus musculu	927	6	0.7	485	1	PTDA_ECOLI	P24241 escherichia
855	6	0.7	448	1	FBL5_RAT	Q9wvh8 rattus norv	928	6	0.7	486	1	ABB3_HUMAN	O95704 homo sapien
856	6	0.7	449	1	YKA3_CAEEL	P34255 caenorhabdi	929	6	0.7	486	1	PRL1_ARATH	Q42384 arabidopsis
857	6	0.7	449	1	HISX_MYCLE	O9cc57 mycobacteri	930	6	0.7	487	1	ENGA_CHLPN	O92762 chlamydia p
858	6	0.7	449	1	TL40_SPTOL	O49939 sphincta ol	931	6	0.7	488	1	FNMG_BRAJA	P28894 bradyrhizob
859	6	0.7	449	1	YBB9_SCHPO	O60071 schizosacch	932	6	0.7	488	1	ENGA_CHLMU	O9pln3 chlamydia m
860	6	0.7	450	1	SYSC_SCHPO	O14018 schizosacch	933	6	0.7	490	1	ENGA_CHLTR	O84709 chlamydia t
861	6	0.7	451	1	HXK_SCHMA	Q26609 schistosoma	934	6	0.7	490	1	GATP_MORCA	Q49092 moraxella c
862	6	0.7	451	1	Y4B3_MYCTU	O11149 mycobacteri	935	6	0.7	490	1	NMT_CRYNE	O25916 helicobacte
863	6	0.7	452	1	SELA_AOUAE	O67140 aquifex aeo	936	6	0.7	491	1	STK3_HUMAN	O13188 homo sapien
864	6	0.7	454	1	TMS3_HUMAN	P57727 homo sapien	937	6	0.7	491	1	CAT1_MAIZE	P18122 zea mays (m
865	6	0.7	455	1	VNS1_BMDNV	O50592 mycobacteri	938	6	0.7	492	1	IMD1_SRRPY	P50099 streptococc
866	6	0.7	455	1	Y142_MYCTU	P48590 drosophila	939	6	0.7	492	1	VSM1_TRYBB	P26331 trypanosoma
867	6	0.7	456	1	HMFT_DROHY	O9bjn1 chlamydia m	940	6	0.7	493	1	CPA9_MESAU	P24453 mesocricetu
868	6	0.7	457	1	SECY_CHIMU	P28639 chlamydia t	941	6	0.7	493	1	GATA_RICCN	O92175 rickettsia
869	6	0.7	457	1	CH1A_YEAST	P38907 saccharomyc	942	6	0.7	493	1	GATA_RICPR	O9ze10 rickettsia
870	6	0.7	458	1	GA45_HUMAN	P31644 homo sapien	943	6	0.7	493	1	NPXR_MOUSE	O99185 mus musculu
871	6	0.7	462	1	COAT_FRG3V	O67473 frog virus	944	6	0.7	494	1	NPXR_RAT	O35764 rattus norv
872	6	0.7	463	1	COAT_FRG3V	O91428 rana tigrin	945	6	0.7	494	1	STK_BACST	O9rhv9 bacillus st
873	6	0.7	463	1	EP11_XENLA	P17506 xenopus lae	946	6	0.7	494	1	AMYB_SOYBN	P10538 glycine max
874	6	0.7	463	1	EP11_XENLA	O01682 schizosacch	947	6	0.7	495	1	GSHR_ORYSA	P48642 oryza sativ
875	6	0.7	463	1	YHCL_BACSU	P54596 bacillus su	948	6	0.7	496	1	HUTH_THEAC	O9h116 thermoplasm
876	6	0.7	463	1	AAKI_PIG	O9wmp4 sus scrofa	949	6	0.7	496	1	TMM_ARATH	O9sdsl arabidopsis
877	6	0.7	464	1	GA45_RAT	P19969 rattus norv	950	6	0.7	497	1	MEK1_YEAST	P24719 saccharomyc
878	6	0.7	465	1	COBB_BACHD	O9kpm8 bacillus ha	951	6	0.7	497	1	CALR_PIG	P25117 sus scrofa
879	6	0.7	466	1	ARLY_AGR75	O8uxx6 agrobacteri	952	6	0.7	498	1	XISC_ANASP	O44217 anabena sp
880	6	0.7	466	1	ARLY_BRUME	O8y1j7 bruceella me	953	6	0.7	498	1	NPXR_HUMAN	O95502 homo sapien
881	6	0.7	466	1	CRD1_COLLI	O01592 columba liv	954	6	0.7	499	1	KPEL_DROME	O95652 drosophila
882	6	0.7	467	1	ATPX_BACFI	P25075 bacillus fi	955	6	0.7	501	1	KDC2_DROME	P16912 drosophila
883	6	0.7	467	1	EUTE_ECOLI	P77445 escherichia	956	6	0.7	502	1	FUT4_DROME	O9vuv9 drosophila
884	6	0.7	467	1	EUTE_ECOLI	P41793 salmonella	957	6	0.7	503	1	TCPT_VIBCH	P29480 vibrio chol
885	6	0.7	468	1	GAG_MSVMY	P32594 moloney mur	958	6	0.7	503	1	CP32_RAT	P05183 rattus norv
886	6	0.7	468	1	NIFB_KLEPN	P10390 klebsiella	959	6	0.7	504	1	CATL_METBA	O93662 methanosarc
887	6	0.7	468	1	PERI_RAT	P21607 rattus norv	960	6	0.7	505	1	EX7L_CAUCR	Q9a649 caulobacter
888	6	0.7	469	1	APPE_DROME	P26673 drosophila	961	6	0.7	505	1	FIXL_BRAJA	P23222 bradyrhizob
889	6	0.7	469	1	KICX_HUMAN	P08779 homo sapien	962	6	0.7	505	1	ATPA_ANASP	P12405 anabena sp
890	6	0.7	470	1	ATPB_BACFI	P22478 bacillus fi	963	6	0.7	506	1	AGP4_MOUSE	O92c00 rickettsia
891	6	0.7	470	1	MTR_NEUCR	P38680 neurospora	964	6	0.7	508	1	SYM_RICPR	O9wv16 mus musculu
892	6	0.7	470	1	RHBA_RHIME	O92312 r diamiohu	965	6	0.7	509	1	AVR1_BOVIN	Q28041 bos taurus
893	6	0.7	470	1	RHBA_RHIME	P41219 homo sapien	966	6	0.7	509	1	AVR1_HUMAN	O04771 homo sapien
894	6	0.7	471	1	YLS7_CAEEL	O9h243 pseudomonas	967	6	0.7	509	1	AVR1_MOUSE	P37172 mus musculu
895	6	0.7	471	1	OTSA_SALTY	P34437 caenorhabdi	968	6	0.7	509	1	AVR1_MOUSE	P48887 rattus norv
896	6	0.7	472	1	XISA_ANASP	O91883 salmonella	969	6	0.7	509	1	COX1_ALBCO	P48887 albiniaria c
897	6	0.7	472	1	DLDH_SYNY3	P72740 synochocyst	970	6	0.7	509	1	ATPA_RHORU	P05036 rhodospill
898	6	0.7	473	1	YCLC_BACSU	O94405 bacillus su	971	6	0.7	510	1	NTP56_DROME	Q27333 drosophila
899	6	0.7	474	1	LEU2_PSEAE	O09645 homo sapien	972	6	0.7	510	1	PUSH_ARATH	Q22928 arabidopsis
900	6	0.7	474	1	SOX4_HUMAN	P77272 escherichia	973	6	0.7	511	1	COPD_BOVIN	P53619 bos taurus
901	6	0.7	474	1	YFEY_ECOLI	Q52964 rhizobium m	974	6	0.7	511	1	ER24_SEPLY	O13597 septoria ly
902	6	0.7	475	1	MOTD_RHIME	P15331 mus musculu	975	6	0.7	512	1	PPX_SALTY	O86091 salmonella
903	6	0.7	475	1	PERI_MOUSE	O92nc6 clostridium	976	6	0.7	512	1	GNTR_BACLI	P46834 bacillus li
904	6	0.7	476	1	CLC_CLOPE	O92nc6 clostridium	977	6	0.7	513	1	GNTR_BACLI	P12011 bacillus su
905	6	0.7	476	1	KPYK_HAEIN	P43994 haemophilus	978	6	0.7	513	1	SR54_STRMG	O54431 streptococc
906	6	0.7	477	1	CATL_LACSK	P30265 lactobacilli	979	6	0.7	516	1	CP53_ASPNG	P17549 aspergillus
907	6	0.7	478	1	DISR_AGRKH	P28364 euplotes oc	980	6	0.7	517	1	CP5H_CANTR	P30610 candida tro
908	6	0.7	478	1	DISR_AGRKH	P28364 euplotes oc	981	6	0.7	517	1	CP5H_CANTR	P30610 candida tro
909	6	0.7	478	1	RPB1_EUPOC	P28364 euplotes oc	982	6	0.7	517	1	CP5H_CANTR	P30610 candida tro

691	6	0.7	397	1	DP3B_MYCSM	P52851 mycobacteri	764	5	0.7	417	1	PVR_HUMAN	P15151 homo sapien
692	6	0.7	397	1	GDN_MOUSE	O07235 mus musculu	765	6	0.7	417	1	SYS_MYCGE	P47251 mycoplasma
693	6	0.7	397	1	GNR_RAT	P07092 rattus norv	766	6	0.7	417	1	URAA_PASMU	O9cp19 pasteurella
694	6	0.7	397	1	PGK_AQUAE	O66519 aquilex aeo	767	6	0.7	418	1	TMW1_MORSP	P11408 moraxella s
695	6	0.7	398	1	AMP1_ARATH	O9s1n5 arabidopsi	768	6	0.7	418	1	RHAA_YERPE	P56507 yersinia pe
696	6	0.7	398	1	KBL_SALTY	P37419 salmonella	769	6	0.7	419	1	CPXY_SYNPY	O02174 yersinia pe
697	6	0.7	399	1	DP3B_MYCLE	P46387 mycobacteri	770	6	0.7	419	1	RHAA_ECOS7	P58506 escherichia
698	6	0.7	399	1	HMPA_BACSU	P49852 bacillus su	771	6	0.7	419	1	RHAA_ECOS7	P58506 escherichia
699	6	0.7	400	1	CSD_CHLMU	O9p1p0 chlamydia m	772	6	0.7	419	1	RHAA_ECOS7	P58506 escherichia
700	6	0.7	400	1	DH12_RAT	P50233 rattus norv	773	6	0.7	420	1	YHAP_BACSU	O07523 bacillus su
701	6	0.7	400	1	ISDE_AGR75	O8uf14 a ispd/ispf	774	6	0.7	420	1	PRE2_STAMM	O94131 gibberella
702	6	0.7	400	1	KICS_HUMAN	O08727 homo sapien	775	6	0.7	420	1	YHAP_BACSU	P22490 streptococ
703	6	0.7	400	1	METK_THEAC	O9hm12 thermoplas	776	6	0.7	421	1	PRE2_HUMAN	P47753 homo sapien
704	6	0.7	400	1	TCR8_PASMU	P51564 pasteurella	777	6	0.7	422	1	TMW1_HUMAN	O75648 homo sapien
705	6	0.7	400	1	YHIN_ECOLI	P37631 escherichia t	778	6	0.7	422	1	MORA_HELPY	P56189 helicobacte
706	6	0.7	401	1	CSD_CHLTR	O84693 chlamydia t	779	6	0.7	423	1	YHAP_HELPY	P30277 rattus norv
707	6	0.7	401	1	O002_RICPR	O9ydy4 aeropyrum p	780	6	0.7	424	1	YHAP_HELPY	P30277 rattus norv
708	6	0.7	401	1	TMW1_AERPE	O9ydy4 aeropyrum p	781	6	0.7	425	1	GAG_PSVGA	O29663 archaeglob
709	6	0.7	402	1	APL3_HUMAN	O95236 homo sapien	782	6	0.7	425	1	SAHH_ANASP	O03337 feline sarc
710	6	0.7	402	1	CPXF_STRGO	P18327 streptococ	783	6	0.7	426	1	ASD4_NEUCR	O8yx05 anabaena sp
711	6	0.7	402	1	DEF_SYNY3	P73881 synecocyst	784	6	0.7	426	1	HEM1_XANCH	O9nevs neurospora
712	6	0.7	402	1	U120_HSVNG	O05101 marek's dis	785	6	0.7	427	1	CREA_ASPPG	O56520 aspergillus
713	6	0.7	402	1	YIN2_STRAM	P32426 streptomyc	786	6	0.7	427	1	DH12_SHEEP	P50168 ovis aries
714	6	0.7	403	1	IF4A_LEIBR	O25225 leishmania	787	6	0.7	427	1	THD1_MYCLE	O94771 mycobacteri
715	6	0.7	403	1	PR56_DICD1	P34123 dictyosteli	788	6	0.7	428	1	DHOM_LACIC	P52985 xanthomonas
716	6	0.7	403	1	RSBP_BACSU	O07014 bacillus su	789	6	0.7	428	1	GBAI_HUMAN	P42808 xanthomonas
717	6	0.7	404	1	SYH_AQUAE	O66522 aquilex aeo	790	6	0.7	428	1	SYS_PASMU	O09302 pasteurella
718	6	0.7	404	1	HCYG_HELPPO	P36823 helix pomat	791	6	0.7	428	1	GBAI_CANAL	P57836 candida alb
719	6	0.7	405	1	DH12_HUMAN	P80365 homo sapien	792	6	0.7	428	1	SYS_PASMU	P57836 candida alb
720	6	0.7	406	1	DH12_RABIT	P51976 oryctolagus	793	6	0.7	429	1	YOP1_CAREL	O09302 pasteurella
721	6	0.7	406	1	HIS2_DEIRA	O9uea3 deinococcus	794	6	0.7	429	1	AGAL_HUMAN	P06280 homo sapien
722	6	0.7	406	1	PEPT_CLOPE	O8xp48 clostridium	795	6	0.7	429	1	CGBI_CRILO	O03280 homo sapien
723	6	0.7	406	1	KCX9_EUGGR	P05729 euglena gra	796	6	0.7	429	1	CGBI_MESAU	O03280 homo sapien
724	6	0.7	407	1	AK2_BACST	P35353 bacillus st	797	6	0.7	429	1	CREA_ASPPG	P37882 mesocricetu
725	6	0.7	408	1	AATC_CAREL	O22067 caenorhabdi	798	6	0.7	429	1	ENO_MOUSE	P41969 mus musculu
726	6	0.7	408	1	PEPT_CLOAB	O97158 clostridium	799	6	0.7	429	1	ENO_MOUSE	O41969 mus musculu
727	6	0.7	408	1	PEPT_ECO57	O8x777 escherichia	800	6	0.7	429	1	SYX_HAEIN	O58992 methanococc
728	6	0.7	408	1	PEPT_ECOLI	P29745 escherichia	801	6	0.7	429	1	VG4_BP1F1	P43833 haemophilus
729	6	0.7	408	1	PEPT_STRAM	O99xv1 staphylococ	802	6	0.7	429	1	YDIS_ECOLI	O03030 bacterioph
730	6	0.7	408	1	THIK_CANTR	P33290 candida tro	803	6	0.7	429	1	CREI_NEUCR	P77337 escherichia
731	6	0.7	408	1	THIK_CANTR	P33291 candida tro	804	6	0.7	430	1	G6P1_MYCPN	O59958 neurospora
732	6	0.7	409	1	HEM1_BRADA	P08262 bradyrhizob	805	6	0.7	430	1	PURA_XYFLA	P78033 mycoplasma
733	6	0.7	409	1	HEM1_BRADA	O82716 salmonella	806	6	0.7	431	1	ENO_ECOLI	O9pg47 xyella flas
734	6	0.7	409	1	PEPT_SALTI	P26311 salmonella	807	6	0.7	431	1	ENO_SALTY	O8324 escherichia
735	6	0.7	410	1	NUOH_MYCTU	P95114 mycobacteri	808	6	0.7	431	1	ENO_YERPE	O82xb2 yersinia pe
736	6	0.7	410	1	PEPT_LISIN	O92am8 listeria in	809	6	0.7	431	1	HISX_LACIA	O02136 lactococcu
737	6	0.7	410	1	PEPT_LISMO	O8f6f0 yersinia pe	810	6	0.7	431	1	NRKA_TRYBB	O08942 trypanosoma
738	6	0.7	411	1	PEPT_VIBCH	O8f6f0 yersinia pe	811	6	0.7	432	1	NRKA_TRYBB	O03428 trypanosoma
739	6	0.7	411	1	LCYB_SYNP7	O9kmy6 vibrio chol	812	6	0.7	432	1	AROA_HAESO	P32310 haemophilus
740	6	0.7	411	1	PEPT_YERPE	O55276 synecococc	813	6	0.7	432	1	YK27_CAREL	O16666 caenorhabdi
741	6	0.7	412	1	DHNO_AGRVI	O44524 agrobacteri	814	6	0.7	433	1	CGBI_HUMAN	P16666 caenorhabdi
742	6	0.7	413	1	DHNO_AGR7	P00366 agrobacteri	815	6	0.7	433	1	CISY_CHICK	P23007 gallus gall
743	6	0.7	413	1	HEMO_MANSE	P31398 manduca sex	816	6	0.7	433	1	ENO_PASMU	P57975 pasteurella
744	6	0.7	413	1	YGS0_YEAST	P45819 saccharomyc	817	6	0.7	433	1	ENO_PASMU	O9kps5 vibrio chol
745	6	0.7	414	1	YGS0_YEAST	O60101 schistosach	818	6	0.7	433	1	ENO_VITRCH	P08778 xenopus lae
746	6	0.7	414	1	YGS0_YEAST	P45819 saccharomyc	819	6	0.7	433	1	KICS_XENTLA	P07955 methanosarc
747	6	0.7	415	1	YGS0_YEAST	O01604 drosophila	820	6	0.7	433	1	Y384_MYCPN	P47624 mycoplasma
748	6	0.7	415	1	YGS0_YEAST	O91599 fowlpox vir	821	6	0.7	433	1	Y384_MYCPN	P75215 mycoplasma
749	6	0.7	415	1	YGS0_YEAST	P39727 saccharomyc	822	6	0.7	434	1	Y384_MYCPN	O94130 boltytis cl
750	6	0.7	416	1	YGS0_YEAST	O01981 emeritella	823	6	0.7	435	1	CREA_BOTCI	O94130 boltytis cl
751	6	0.7	416	1	YGS0_YEAST	P52993 alcaligenes	824	6	0.7	435	1	ENO_BOTCI	O94130 boltytis cl
752	6	0.7	416	1	YGS0_YEAST	P29408 macropus eu	825	6	0.7	435	1	ENO_BOTCI	O94130 boltytis cl
753	6	0.7	416	1	YGS0_YEAST	P09411 mus musculu	826	6	0.7	435	1	ENO_BOTCI	O94130 boltytis cl
754	6	0.7	416	1	YGS0_YEAST	P16617 rattus norv	827	6	0.7	435	1	ENO_BOTCI	O94130 boltytis cl
755	6	0.7	416	1	YGS0_YEAST	P50310 equus caball	828	6	0.7	435	1	ENO_BOTCI	O94130 boltytis cl
756	6	0.7	416	1	YGS0_YEAST	P00559 equus caball	829	6	0.7	435	1	ENO_BOTCI	O94130 boltytis cl
757	6	0.7	416	1	YGS0_YEAST	P41759 schistosoma	830	6	0.7	435	1	ENO_BOTCI	O94130 boltytis cl
758	6	0.7	417	1	YGS0_YEAST	O9c884 lactococcus	831	6	0.7	437	1	INO_MYCTU	O06628 mycobacteri
759	6	0.7	417	1	YGS0_YEAST	P00558 homo sapien	832	6	0.7	437	1	INO_MYCTU	O06628 mycobacteri
760	6	0.7	417	1	YGS0_YEAST	P91427 caenorhabdi	833	6	0.7	437	1	INO_MYCTU	O06628 mycobacteri
761	6	0.7	417	1	YGS0_YEAST	P46727 caenorhabdi	834	6	0.7	438	1	INO_MYCTU	O06628 mycobacteri
762	6	0.7	417	1	YGS0_YEAST	P41157 candida mal	835	6	0.7	439	1	INO_MYCTU	O06628 mycobacteri
763	6	0.7	417	1	YGS0_YEAST	P32506 cercopithec	836	6	0.7	439	1	INO_MYCTU	O06628 mycobacteri

545	6	0.7	342	1	GCP_HAEIN	P43764	haemophilus	618	6	0.7	369	1	ADH3_ECOLI	P25437	escherichia
546	6	0.7	342	1	QUEA_BACSU	032054	bacillus su	619	6	0.7	369	1	ADH3_PASPI	P39450	pasteurella
547	6	0.7	342	1	RTCA_TYRAB	090206	pyrococcus	620	6	0.7	369	1	FENR_SPILO	P09455	spiraclea ol
548	6	0.7	342	1	YPTX_CLOPE	P18020	clostridium	621	6	0.7	369	1	YHAT_LPMV	P27287	ja.piedad-m
549	6	0.7	343	1	HOLA_ECOLI	P28630	escherichia	622	6	0.7	369	1	Y736_HELPY	025436	helicobacte
550	6	0.7	345	1	CC22_TRYBB	P54655	trypanosoma	623	6	0.7	370	1	FLGI_CAUCR	P33979	caulobacter
551	6	0.7	345	1	YLX4_CAEEL	P46500	caenorhabdi	624	6	0.7	370	1	PSPB_RABIT	P15285	oryctolagus
552	6	0.7	345	1	YPR1_AERHY	005485	aeromonas h	625	6	0.7	370	1	TFPD_ALCEU	P05404	alcaligenes
553	6	0.7	346	1	YF77_MYCRN	P75203	mycoplasma	626	6	0.7	371	1	GAG_FSVST	P03388	feline sarc
554	6	0.7	347	1	DLX4_HUMAN	092888	homo sapien	627	6	0.7	371	1	YOJE_BACSU	P54542	bacillus su
555	6	0.7	347	1	MU2M_HALGR	P38599	hallicoeris	628	6	0.7	374	1	PILT_NETGO	006581	neisseria g
556	6	0.7	347	1	MU2M_PHOVI	000540	phoca vitul	629	6	0.7	374	1	TGT_ANASP	Q8HYVC	anaeadaea sp
557	6	0.7	347	1	NUSA_MYCTU	005618	mycobacteri	630	6	0.7	375	1	ADHR_GADMO	P81600	gadus morhu
558	6	0.7	348	1	Y001_ARCFU	030701	archaeoglob	631	6	0.7	375	1	MASP_HUMAN	P36592	homo sapien
559	6	0.7	350	1	ENP2_TORCA	P14401	torpeda cal	632	6	0.7	375	1	MASP_RAT	P70564	rattus norv
560	6	0.7	350	1	STFA_THERH	P27001	thermus the	633	6	0.7	376	1	AADE_YEAST	P42884	saccharomyc
561	6	0.7	351	1	NOV_CHICK	P28686	gallus gall	634	6	0.7	376	1	ADHX_MYXGL	P80362	myxine glut
562	6	0.7	351	1	ROM1_MOUSE	P32958	mus muscull	635	6	0.7	376	1	G4PI_YEAST	P46630	saccharomyc
563	6	0.7	352	1	H218_RAT	P47752	rattus norv	636	6	0.7	376	1	METR_MYCPU	Q98PM0	mycoplasma
564	6	0.7	352	1	HEM2_BRUME	093692	bruceella me	637	6	0.7	376	1	PYRE_CEPAC	P14017	cephalospor
565	6	0.7	352	1	TH14_ALINGL	038709	alnus glut	638	6	0.7	378	1	CARA_PSEAE	P38098	pseudomonas
566	6	0.7	353	1	NOV_COTJA	P42542	colutrinx co	639	6	0.7	378	1	PYRX_HELPY	025045	helicobacte
567	6	0.7	353	1	PROB_THEMA	09WYD0	thermotoga	640	6	0.7	379	1	CYB_SPEMO	Q9TF31	spermophilu
568	6	0.7	353	1	YCJF_ECOLI	P45525	escherichia	641	6	0.7	379	1	HS7X_PIG	P49424	sus scrofa
569	6	0.7	353	1	YPHC_ECOLI	P77360	escherichia	642	6	0.7	379	1	Y4XO_RHISN	P55707	rhizobium s
570	6	0.7	354	1	ARC2_HUMAN	P78540	homo sapien	643	6	0.7	380	1	RAD2_SCHPO	P39750	schizosacch
571	6	0.7	354	1	CARC_HUMAN	043570	homo sapien	644	6	0.7	380	1	SYW_HALNI	Q9HN66	halobacteri
572	6	0.7	354	1	CARX_BACST	P54324	bacillus st	645	6	0.7	381	1	EDGI_HUMAN	P21453	homo sapien
573	6	0.7	354	1	GRI_HOMAM	P41776	homarus ame	646	6	0.7	382	1	EDGI_MOUSE	008530	mus muscull
574	6	0.7	354	1	NOV_MOUSE	Q64299	mus muscull	647	6	0.7	382	1	RSIH_BACSU	P38454	bacillus su
575	6	0.7	355	1	ARGC_AERPE	Q9Y9H8	aeropyrum p	648	6	0.7	382	1	YI03_CAEEL	001811	caenorhabdi
576	6	0.7	355	1	C3X1_HUMAN	P49338	homo sapien	649	6	0.7	383	1	EDGI_RAT	P48303	rattus norv
577	6	0.7	355	1	FKB8_HUMAN	Q14318	homo sapien	650	6	0.7	383	1	LA_AEDAL	Q26457	aedes albop
578	6	0.7	355	1	VEA1_NPVOP	Q35465	mus muscull	651	6	0.7	383	1	QUEA_RICPR	09ZDE6	ricchettiata
579	6	0.7	355	1	HIS7_VIBCH	010278	orygia pseu	652	6	0.7	384	1	MURG_TREPA	083535	treponema p
580	6	0.7	357	1	ISPQ_CAMJE	Q9KXK1	vibriol chol	653	6	0.7	384	1	PRTK_TRIAL	P39771	trichachiu
581	6	0.7	357	1	LDGX_MALDO	Q9P6M1	campylobact	654	6	0.7	384	1	PURF_BACSU	P39771	bacillus su
582	6	0.7	357	1	NOV_SAPIEN	P51091	malus domes	655	6	0.7	385	1	CYB_MYXGL	Q09240	myxine glut
583	6	0.7	357	1	ACOD_RAT	P97308	rattus norv	656	6	0.7	385	1	GBA5_CAEEL	Q20701	caenorhabdi
584	6	0.7	358	1	AROB_CORPS	P67449	corynebacte	657	6	0.7	385	1	METR_HELPY	019383	mycobacteri
585	6	0.7	358	1	LIP_PSEGL	Q05489	pseudomonas	658	6	0.7	385	1	YMBM_BACSU	P39566	bacillus su
586	6	0.7	358	1	PIAP_PIG	Q62640	sus scrofa	659	6	0.7	385	1	CRTI_ERWHE	001331	erwinia her
587	6	0.7	359	1	AG2R_CHICK	P79785	gallus gall	660	6	0.7	386	1	PPAP_HUMAN	P15309	homo sapien
588	6	0.7	359	1	AC2R_MELGA	P33396	meleagris g	661	6	0.7	386	1	SUCC_COXBU	P53592	coxellella bu
589	6	0.7	359	1	CARX_BACHD	Q9KXV6	bacillus ha	662	6	0.7	386	1	TIAL_HUMAN	P31483	homo sapien
590	6	0.7	359	1	CGIP_SCHPO	P25009	schizosacch	663	6	0.7	386	1	TIAL_MOUSE	P52912	mus muscull
591	6	0.7	359	1	ISPG_HELPY	025342	helicobacte	664	6	0.7	387	1	OMAO_YEAST	P23644	saccharomyc
592	6	0.7	359	1	ISPG_HELPY	025342	helicobacte	665	6	0.7	387	1	YGH5_YEAST	P53149	saccharomyc
593	6	0.7	359	1	MAS2_AGRRH	P27260	tomato yell	666	6	0.7	387	1	ALR_MYCAV	Q91888	mycobacteri
594	6	0.7	359	1	VAL1_TYLCM	007733	mycobacteri	667	6	0.7	388	1	PYRD_DOMOE	P32748	drosophila
595	6	0.7	359	1	Y199_MYCTU	P58894	streptococc	668	6	0.7	389	1	SERI_BOMMO	P07848	domomyx mori
596	6	0.7	360	1	CARA_STRPB	Q9A0C7	streptococc	669	6	0.7	389	1	MTRK_METEX	P53594	methyllobact
597	6	0.7	360	1	CARA_STRPB	Q9A0C7	streptococc	670	6	0.7	390	1	CARA_XANAC	P58895	xanthomonas
598	6	0.7	360	1	MDH_PYRHO	059028	pyrococcus	671	6	0.7	391	1	YIDJ_ECOLI	Q9U48	homo sapien
599	6	0.7	360	1	PSBA_GALSU	P24725	galdieria s	672	6	0.7	391	1	YIDJ_ECOLI	P31462	escherichia
600	6	0.7	361	1	LEU1_PYRHO	059390	pyrococcus	673	6	0.7	392	1	DSOR_DOMOE	P07062	flavobacter
601	6	0.7	361	1	PAX1_HUMAN	P15863	homo sapien	674	6	0.7	392	1	THIL_ALCEU	Q24334	drosophila
602	6	0.7	361	1	PAX1_MOUSE	P09084	mus muscull	675	6	0.7	393	1	YEAN_ECOLI	P14611	alcaligenes
603	6	0.7	362	1	CHSA_IPOPL	P48400	ipomoea pia	676	6	0.7	393	1	AROC_STRCO	P76242	alcaligenes
604	6	0.7	362	1	SRK2_SPOLA	P42888	spongilla l	677	6	0.7	394	1	EUTE_ECOLI	Q9KXG4	streptomyce
605	6	0.7	363	1	CTF1_SCHPO	043040	schizosacch	678	6	0.7	395	1	RRP_NNVU2	P76553	escherichia
606	6	0.7	363	1	MURG_BACSU	P37585	bacillus su	679	6	0.7	395	1	RRP_NNVU2	P41795	salmonella
607	6	0.7	364	1	YK27_AQUAE	Q67821	pseudomonas	680	6	0.7	395	1	ARRS_RANCA	P16073	newcastle d
608	6	0.7	364	1	GCST_BACHD	Q9KX34	baacillus ha	681	6	0.7	396	1	DH12_MOUSE	Q06427	newcastle d
609	6	0.7	365	1	MURA_ECOLI	P46885	escherichia	682	6	0.7	396	1	ARRS_RANCA	P51478	rana catesb
610	6	0.7	365	1	PARB_MOUSE	Q9E646	mus muscull	683	6	0.7	396	1	DXR_XYLFA	P51661	mus muscull
611	6	0.7	367	1	FTSZ_AQUAE	060931	homo sapien	684	6	0.7	396	1	LIDD_ECOLI	Q9PE10	xyellella fas
612	6	0.7	367	1	TRMU_MYCGE	P47537	mycoplasma	685	6	0.7	396	1	PRTY_BOVIN	P10349	cucurbita m
613	6	0.7	368	1	GUN_ECO57	08X519	escherichia	686	6	0.7	396	1	YNE2_YEAST	P53958	saccharomyc
614	6	0.7	368	1	GUN_ECOLI	P37651	escherichia	687	6	0.7	396	1			
615	6	0.7	368	1	GUN_ECOLI	P37651	escherichia	688	6	0.7	396	1			
616	6	0.7	368	1	GUN_ECOLI	P37651	escherichia	689	6	0.7	396	1			
617	6	0.7	368	1	GUN_ECOLI	P37651	escherichia	690	6	0.7	396	1			

399	6	0.7	281	1	ARAQ_SALTY	P03022 salmonella	472	0.7	313	1	YK89_ARCFU	O28190 archaeoglob
400	6	0.7	281	1	PANC_XYLF	O969r xyella fas	473	0.7	314	1	ETFA_BRAJA	P53573 bradyrhizob
401	6	0.7	281	1	Y125_MYCP	P75511 mycoplasma	474	0.7	314	1	KHSE_HAEIN	P44504 haemophilus
402	6	0.7	282	1	RIP2_BRVDI	P98184 dryonia dio	475	0.7	314	1	KHSE_PASMU	O94030 pasteurella
403	6	0.7	282	1	Y134_MERYA	O57598 methanococ	476	0.7	315	1	SPSI_BACSU	P39630 bacillus su
404	6	0.7	283	1	GAG1_HUMAN	P10263 homo sapien	477	0.7	316	1	COYT_YEAST	P53339 saccharomyc
405	6	0.7	283	1	HOXQ_AZOVI	P40594 azotobacter	478	0.7	316	1	GSEP_BACLI	P80057 bacillus li
406	6	0.7	283	1	TONB_NEIGO	O06432 neisseria g	479	0.7	317	1	HEW2_DEIRA	O94998 deinococcus
407	6	0.7	284	1	HUPQ_AZOC	O43957 azotobacter	480	0.7	317	1	HEW3_CIAL	P34797 clarkia lew
408	6	0.7	285	1	FLA3_TREMA	O94998 treponema m	481	0.7	317	1	YVCK_BACSU	O94998 mycobacteri
409	6	0.7	286	1	DAPE_BACHD	O94710 bacillus ha	482	0.7	318	1	KHSE_VIBCH	O94998 mycobacteri
410	6	0.7	286	1	PAL_KLEPN	P37446 klebsiella	483	0.7	319	1	HUTG_BACHD	O94998 mycobacteri
411	6	0.7	286	1	PARB_RICPR	O94710 bacillus ha	484	0.7	319	1	PTFC_MYCLE	O94998 mycobacteri
412	6	0.7	286	1	RER2_YEAST	P35196 saccharomyc	485	0.7	319	1	TAL_RALSO	O94998 mycobacteri
413	6	0.7	287	1	Y791_CHLMU	O94998 chlamydia m	486	0.7	320	1	HLPI_HUMAN	O94998 mycobacteri
414	6	0.7	287	1	YICC_ECOLI	P23839 escherichia	487	0.7	321	1	LPD_CAME	O94998 mycobacteri
415	6	0.7	287	1	YICC_HAEIN	P44726 haemophilus	488	0.7	321	1	NECD_HUMAN	O94998 mycobacteri
416	6	0.7	288	1	AADA_YEAST	P47182 saccharomyc	489	0.7	321	1	YACU_RHISN	O94998 mycobacteri
417	6	0.7	288	1	Y504_CHLTR	O84512 chlamydia t	490	0.7	321	1	YACU_RHISN	O94998 mycobacteri
418	6	0.7	289	1	THTR_CHICK	P23324 gallus gall	491	0.7	321	1	YACU_RHISN	O94998 mycobacteri
419	6	0.7	291	1	RGR_BOVIN	P47803 bos taurus	492	0.7	322	1	YVBD_BACSU	O94998 mycobacteri
420	6	0.7	291	1	RGR_MOUSE	O94998 mus musculu	493	0.7	322	1	YVBD_BACSU	O94998 mycobacteri
421	6	0.7	292	1	ARAQ_ECOLI	P03021 escherichia	494	0.7	323	1	CC39_CAEEL	O94998 mycobacteri
422	6	0.7	292	1	E434_ADE09	P89083 human adeno	495	0.7	323	1	FRG1_CAEEL	O94998 mycobacteri
423	6	0.7	292	1	KHSE_PYRPU	O84129 pyrococcus	496	0.7	323	1	SECF_ECOLI	O94998 mycobacteri
424	6	0.7	293	1	RCL_CAEEL	P49405 caenorhabdi	497	0.7	325	1	YRBF_ECOLI	O94998 mycobacteri
425	6	0.7	293	1	ROCL_HUMAN	O60812 homo sapien	498	0.7	326	1	BRUC_ECO57	O94998 mycobacteri
426	6	0.7	294	1	E434_ADE02	P03239 human adeno	499	0.7	326	1	BRUC_ECO57	O94998 mycobacteri
427	6	0.7	294	1	NPM_CHICK	P16039 gallus gall	500	0.7	326	1	BRUC_ECO57	O94998 mycobacteri
428	6	0.7	295	1	RPSQ_MYXXA	O07083 myxococcus	501	0.7	326	1	BRUC_ECO57	O94998 mycobacteri
429	6	0.7	295	1	THTR_RAT	P24329 rattus norv	502	0.7	326	1	BRUC_ECO57	O94998 mycobacteri
430	6	0.7	296	1	DAPE_MYCLE	P46814 mycobacteri	503	0.7	327	1	ODPB_RICPR	O94998 mycobacteri
431	6	0.7	296	1	NPL_ECOLI	P06995 escherichia	504	0.7	327	1	DRN1_STREQ	O94998 mycobacteri
432	6	0.7	296	1	THTR_BOVIN	P00586 bos taurus	505	0.7	328	1	YGB7_ALCEU	O94998 mycobacteri
433	6	0.7	296	1	THTR_CRIGR	P46635 cricetus	506	0.7	328	1	MOA_RHOCA	O94998 mycobacteri
434	6	0.7	296	1	URER_ECOLI	P23236 escherichia	507	0.7	328	1	OPT_MOUSE	O94998 mycobacteri
435	6	0.7	296	1	Y639_SYNY3	O55717 synechocyst	508	0.7	328	1	YE27_MERJA	O94998 mycobacteri
436	6	0.7	297	1	YK46_METAC	P57880 methanococ	509	0.7	329	1	AAD4_YEAST	O94998 mycobacteri
437	6	0.7	299	1	G3P_DICDI	O94489 dictyostell	510	0.7	330	1	G3PC_TRYBB	O94998 mycobacteri
438	6	0.7	299	1	MRW_THEMA	O94489 dictyostell	511	0.7	330	1	HEM2_MOUSE	O94998 mycobacteri
439	6	0.7	300	1	ARGW_MERJA	O60382 thermotoga	512	0.7	330	1	LAPE_VIRPA	O94998 mycobacteri
440	6	0.7	300	1	JAM1_MOUSE	O60382 methanococ	513	0.7	330	1	LAPE_VIRPA	O94998 mycobacteri
441	6	0.7	301	1	Y1DE_ECOLI	O88792 mus musculu	514	0.7	330	1	YBCB_BACSU	O94998 mycobacteri
442	6	0.7	301	1	Y1DE_ECOLI	P32667 caenorhabdi	515	0.7	331	1	DGAL_HAEIN	O94998 mycobacteri
443	6	0.7	302	1	PTB_CLOBE	O05624 clostridium	516	0.7	331	1	GSPK_AERHY	O94998 mycobacteri
444	6	0.7	303	1	UI24_HSVSA	O01005 herpessvirus	517	0.7	332	1	G3P_PHYTN	O94998 mycobacteri
445	6	0.7	303	1	YB81_YEAST	P38326 saccharomyc	518	0.7	333	1	OTCI_ECOLI	O94998 mycobacteri
446	6	0.7	304	1	MDH_HAUMA	O07841 halocaula	519	0.7	333	1	SYW_PASMU	O94998 mycobacteri
447	6	0.7	305	1	NAC_ECOLI	O47005 escherichia	520	0.7	333	1	T2X1_XANCC	O94998 mycobacteri
448	6	0.7	305	1	PPPE_RAT	O64620 rattus norv	521	0.7	333	1	VRPO_LAMBD	O94998 mycobacteri
449	6	0.7	305	1	Y348_MYCPN	P75255 mycoplasma	522	0.7	333	1	Y395_RICPR	O94998 mycobacteri
450	6	0.7	306	1	PYRH_NEIMA	O94998 neisseria m	523	0.7	333	1	Y4M_RHISN	O94998 mycobacteri
451	6	0.7	306	1	ROC_HUMAN	O94998 neisseria m	524	0.7	334	1	G3P_BACCO	O94998 mycobacteri
452	6	0.7	306	1	YC73_AQUAE	O67310 aquifex aeo	525	0.7	334	1	LCMT_HUMAN	O94998 mycobacteri
453	6	0.7	308	1	FMT_CAUCR	O94998 caulobacter	526	0.7	336	1	KC1B_BOVIN	O94998 mycobacteri
454	6	0.7	308	1	HCGP_MOUSE	O02614 mus musculu	527	0.7	337	1	COBU_PSEDE	O94998 mycobacteri
455	6	0.7	309	1	CD86_MOUSE	O94998 mus musculu	528	0.7	337	1	YEB7_YEAST	O94998 mycobacteri
456	6	0.7	309	1	KHSE_SALTY	O84925 serralia ma	529	0.7	338	1	MOZK_CAEEL	O94998 mycobacteri
457	6	0.7	309	1	KHSE_SALTY	O84925 serralia ma	530	0.7	338	1	OPPD_LACLA	O94998 mycobacteri
458	6	0.7	309	1	KHSE_STRMA	O94998 streptomyces	531	0.7	338	1	OPPD_LACLA	O94998 mycobacteri
459	6	0.7	309	1	KHSE_STRMA	O94998 streptomyces	532	0.7	338	1	OPPD_LACLA	O94998 mycobacteri
460	6	0.7	309	1	KHSE_STRMA	O94998 streptomyces	533	0.7	338	1	OPPD_LACLA	O94998 mycobacteri
461	6	0.7	309	1	KHSE_STRMA	O94998 streptomyces	534	0.7	338	1	OPPD_LACLA	O94998 mycobacteri
462	6	0.7	310	1	KHSE_STRMA	O94998 streptomyces	535	0.7	339	1	OPPD_LACLA	O94998 mycobacteri
463	6	0.7	310	1	KHSE_STRMA	O94998 streptomyces	536	0.7	339	1	OPPD_LACLA	O94998 mycobacteri
464	6	0.7	310	1	KHSE_STRMA	O94998 streptomyces	537	0.7	339	1	OPPD_LACLA	O94998 mycobacteri
465	6	0.7	310	1	KHSE_STRMA	O94998 streptomyces	538	0.7	339	1	OPPD_LACLA	O94998 mycobacteri
466	6	0.7	311	1	KHSE_STRMA	O94998 streptomyces	539	0.7	340	1	OPPD_LACLA	O94998 mycobacteri
467	6	0.7	311	1	KHSE_STRMA	O94998 streptomyces	540	0.7	341	1	OPPD_LACLA	O94998 mycobacteri
468	6	0.7	311	1	KHSE_STRMA	O94998 streptomyces	541	0.7	341	1	OPPD_LACLA	O94998 mycobacteri
469	6	0.7	312	1	KHSE_STRMA	O94998 streptomyces	542	0.7	341	1	OPPD_LACLA	O94998 mycobacteri
470	6	0.7	313	1	KHSE_STRMA	O94998 streptomyces	543	0.7	341	1	OPPD_LACLA	O94998 mycobacteri
471	6	0.7	313	1	KHSE_STRMA	O94998 streptomyces	544	0.7	341	1	OPPD_LACLA	O94998 mycobacteri

253	6	0.7	193	1	IL18_BOVIN	Q9LU73	bos taurus	326	6	0.7	234	1	YV04_MYCTU	O50721	mycobacteri
254	6	0.7	193	1	IL18_CANFA	Q9XER0	canis famill	327	6	0.7	235	1	GAG_MLVAB	P03333	abelson mur
255	6	0.7	193	1	IL18_HORSE	Q9XSG7	equus caball	328	6	0.7	235	1	SFSA_PSEAE	Q9HW77	pseudomonas
256	6	0.7	193	1	ORN_CABEL	O17819	caenorhabdit	329	6	0.7	236	1	CLCD_PSEPU	P11453	pseudomonas
257	6	0.7	194	1	RS9B_YEAST	P05755	saccharomyc	330	6	0.7	236	1	LYDE_LACLA	O9CE00	laccococcus
258	6	0.7	194	1	Y248_METJA	O57698	methanococc	331	6	0.7	238	1	Y181_BUCAI	P57278	buchnera ap
259	6	0.7	195	1	YOHG_ECOLI	P33365	escherichia	332	6	0.7	240	1	KDKA_PASPI	P58551	pasteurella
260	6	0.7	195	1	3MCH_CHLPN	Q9Z847	chlamydia p	333	6	0.7	241	1	COMB_THETN	O8EM03	thermoanaer
261	6	0.7	196	1	AMIR_PSEAE	P10932	pseudomonas	334	6	0.7	241	1	KDNM_BP74	P35113	agrobacteri
262	6	0.7	196	1	BRUR_SALRY	P31575	salmomella	335	6	0.7	241	1	NOCK_AGR5	P40117	alcaligenes
263	6	0.7	196	1	HADB_EURPI	O53009	burkholderi	336	6	0.7	241	1	CYBH_BRAJA	P21960	bradyrhizob
264	6	0.7	196	1	LAST_SERMA	P37006	serratia ma	337	6	0.7	244	1	HISA_MYCTU	O06588	mycobacteri
265	6	0.7	196	1	RNH2_AOUAE	O67768	aquifex aeo	338	6	0.7	244	1	AX1B_ARATH	Q38829	halobacteris
266	6	0.7	196	1	RS9A_YEAST	O13516	saccharomyc	339	6	0.7	246	1	PCNA_HA1N1	Q9HW45	halobacteri
267	6	0.7	198	1	RL19_CABEL	O02639	caenorhabdi	340	6	0.7	247	1	RECO_BRUME	O8Y767	bruceella me
268	6	0.7	198	1	RM12_HUMAN	P52815	homo sapien	341	6	0.7	247	1	VAB7_CABEL	O93899	caenorhabdi
269	6	0.7	200	1	ACPD_SALTY	O8XTP4	salmonella	342	6	0.7	248	1	MIRP_ECOLI	P77486	escherichia
270	6	0.7	200	1	RPOE_STRPN	Q97894	streptococc	343	6	0.7	248	1	RASH_RKASV	P01114	rashed rat
271	6	0.7	200	1	YPT2_SCHPO	P17609	schizosacch	344	6	0.7	248	1	COX2_NEUCR	P00411	neutrospora
272	6	0.7	201	1	CA01_HUMAN	O76042	homo sapien	345	6	0.7	250	1	KDSB_YERPE	O8Z984	yersinia pe
273	6	0.7	201	1	RM12_MOUSE	O9D815	mus musculu	346	6	0.7	250	1	PYRK_STRPN	O97166	streptococc
274	6	0.7	201	1	RIYA_STRCO	O91390	streptomyce	347	6	0.7	250	1	YJ72_CORGL	P40111	corynebacte
275	6	0.7	201	1	YMH9_YEAST	O03629	saccharomyc	348	6	0.7	250	1	YC71_MYCPN	P75505	mycoplasma
276	6	0.7	202	1	HPRT_MYCTU	O06383	mycobacteri	349	6	0.7	251	1	YCA1_PLAFA	P25407	plasmodium
277	6	0.7	202	1	WRB2_RHIME	Q93012	rhizobium m	350	6	0.7	252	1	RECO_RH1LO	Q98543	rhizobium 1
278	6	0.7	202	1	YAJ1_PSEAE	O9HW45	pseudomonas	351	6	0.7	253	1	YD69_MYCPN	P75412	mycoplasma
279	6	0.7	203	1	HPRT_MYCAV	P96794	mycobacteri	352	6	0.7	253	1	YF05_MYCPN	P75281	mycoplasma
280	6	0.7	203	1	HPRT_MYCLE	O69537	mycobacteri	353	6	0.7	253	1	YC1L_ECOLI	P76038	escherichia
281	6	0.7	203	1	KGUA_PSEAE	O9HW45	pseudomonas	354	6	0.7	254	1	CU02_HUMAN	O43822	homo sapien
282	6	0.7	203	1	RM12_CRICR	P52827	cricketus cr	355	6	0.7	256	1	Y480_THHAC	O9HW46	thermoplasm
283	6	0.7	205	1	YETR_SHEON	P46149	shevanella	356	6	0.7	257	1	Y480_HAEIN	P45024	haemophilus
284	6	0.7	206	1	PBS2_TRYBB	O9HW46	trypanosoma	357	6	0.7	257	1	MODD_RHOCA	Q08334	rhodobacter
285	6	0.7	206	1	RS4_BUCAP	P57567	buchnera ap	358	6	0.7	259	1	TPIS_CORGL	P19583	corynebacte
286	6	0.7	206	1	SODM_LACIA	P50911	lactococcus	359	6	0.7	259	1	CCG6_MOUSE	O8VHW7	rattus norv
287	6	0.7	206	1	Y21K_MSVN	P14979	maize strea	360	6	0.7	260	1	Y4H4_RH1SN	P55477	rattus norv
288	6	0.7	206	1	MGMT_HUMAN	P16455	homo sapien	361	6	0.7	260	1	AX1A_ARATH	Q28828	arabidopsis
289	6	0.7	207	1	IF2B_THEVO	Q97805	thermoplasm	362	6	0.7	261	1	COX3_MYXGL	O63915	myxine glut
290	6	0.7	208	1	MGMT_RAT	P24528	rattus norv	363	6	0.7	261	1	TPIS_MYCTU	O08408	mycobacteri
291	6	0.7	208	1	NORD_HAEIN	P43593	haemophilus	364	6	0.7	261	1	YAFI_ECOLI	P77339	escherichia
292	6	0.7	209	1	IF2B_THEAC	O9HW45	thermoplasm	365	6	0.7	261	1	ATP6_HAEIN	P43379	haemophilus
293	6	0.7	209	1	MGMT_CRICR	P26186	cricketus	366	6	0.7	262	1	MOBB_BP4	P13329	bacterioph
294	6	0.7	209	1	URK_DEIRA	O91X45	delinococcus	367	6	0.7	262	1	TRPA_STY3	P77960	synechocyst
295	6	0.7	210	1	DYR_YEAST	O69169	yersinia pe	368	6	0.7	264	1	PYRK_MERTH	O47280	methanobact
296	6	0.7	211	1	KTHY_YERPE	P48293	mus musculu	369	6	0.7	264	1	SUHB_HAEIN	P43333	haemophilus
297	6	0.7	211	1	Y767_MYCTU	P48293	mus musculu	370	6	0.7	267	1	SUHB_PASMU	O59878	pasteurella
298	6	0.7	213	1	ET3_MOUSE	O91498	cyanidium c	371	6	0.7	267	1	ISPE_HELPY	O25984	helicobacte
299	6	0.7	214	1	HIS5_CYACA	O58370	methanococc	372	6	0.7	268	1	PARA_MYCGE	P47706	mycoplasma
300	6	0.7	214	1	TAL_METJA	O89116	mus musculu	373	6	0.7	269	1	Y513_STRAU	O05337	staphylococ
301	6	0.7	217	1	VTL2_MOUSE	O27765	methanobact	374	6	0.7	269	1	CH12_PERTH	P13869	petunia hyb
302	6	0.7	217	1	YH32_METTH	O9UW88	pyrococcus	375	6	0.7	270	1	THIM_BACSD	P39593	bacillus su
303	6	0.7	217	1	RL1_PYRAB	O57782	pyrococcus	376	6	0.7	272	1	GRM_MICBC	O47539	escherichia
304	6	0.7	219	1	PAIP2_CAPVK	P19748	capripoxvir	377	6	0.7	272	1	MCT6_RAT	P53716	mycoplasma
305	6	0.7	220	1	YIEH_ECOLI	O9A616	caulobacter	378	6	0.7	274	1	MORI_BRUME	O8YH42	bruceella me
306	6	0.7	220	1	PIMT_CAUCR	O9A616	caulobacter	379	6	0.7	274	1	Y623_CHLPN	O9Z7E1	chlamydia p
307	6	0.7	221	1	TPIS_METAC	O58400	methanococc	380	6	0.7	277	1	TAUC_ECOLI	P55073	homo sapien
308	6	0.7	222	1	COMB_METJA	O51659	paracoccus	381	6	0.7	278	1	HEMK_ECOLI	O41218	mus musculu
309	6	0.7	222	1	MAUR_PARDE	P50051	mycobacteri	382	6	0.7	278	1	RECO_BRUME	P49897	rattus norv
310	6	0.7	224	1	URB5_MYCTU	O9J151	rattus norv	383	6	0.7	278	1	AX1B_ARATH	P53174	saccharomyc
311	6	0.7	224	1	VTL2_RAT	O58355	methanococc	384	6	0.7	278	1	PCNA_HA1N1	P95374	neisseria m
312	6	0.7	224	1	Y945_METJA	O43097	thermomyc	385	6	0.7	278	1	KDSB_YERPE	P54966	paramecium
313	6	0.7	224	1	XYNA_THELA	P92400	equus asinu	386	6	0.7	278	1	PYRK_STRPN	P57003	neisseria m
314	6	0.7	225	1	ATP6_EQUAS	P05504	rattus norv	387	6	0.7	278	1	YJ72_CORGL	P57004	neisseria m
315	6	0.7	226	1	ATP6_RAT	O01453	bacillus su	388	6	0.7	278	1	YC71_MYCPN	O58622	methanococc
316	6	0.7	226	1	MINC_BACSU	O01453	bacillus su	389	6	0.7	278	1	YCA1_PLAFA	P11765	citrobacter
317	6	0.7	226	1	MINC_MYCPU	O98182	mycoplasma	390	6	0.7	278	1	YD69_MYCPN		
318	6	0.7	226	1	YMO4_CABEL	P34495	caenorhabdi	391	6	0.7	278	1	YF05_MYCPN		
319	6	0.7	227	1	PSBQ_ONOVI	O22591	onoprychis	392	6	0.7	278	1	YAFI_ECOLI		
320	6	0.7	227	1	MINC_BACHD	O9C888	bacillus ha	393	6	0.7	278	1	ATP6_HAEIN		
321	6	0.7	232	1	LPKX_MYCLE	O27214	methanobact	394	6	0.7	278	1	MOBB_BP4		
322	6	0.7	233	1	YB46_METTH			395	6	0.7	278	1	TRPA_STY3		
323	6	0.7	234	1				396	6	0.7	278	1	PYRK_MERTH		
324	6	0.7	234	1				397	6	0.7	278	1	SUHB_HAEIN		
325	6	0.7	234	1				398	6	0.7	278	1	ISPE_HELPY		

107	6	0.7	51	1	AC_BPT4	P18924 bacterioph	180	0.7	149	1	RL19_ARCFU	O28372 archaeoglob
108	6	0.7	52	1	YLM3_CAEFL	P34377 caenornabai	181	0.7	151	1	RL7A_PYPAR	O8245 pyrobaculum
109	6	0.7	54	1	ELHC_APLCA	P11924 aplysia cal	182	0.7	152	1	PSAL_GUTTH	O7865 pylliardia
110	6	0.7	54	1	ELHC_APLCA	P11925 aplysia cal	183	0.7	152	1	Y533_CHLPM	O92819 chlamydia p
111	6	0.7	56	1	H71A_MANSE	P26226 manduca sex	184	0.7	152	1	Y686_CHLPM	O99120 chlamydia m
112	6	0.7	57	1	RS27_HALN1	O9nrt7 halobacteri	185	0.7	152	1	YFE7_YEAST	P43346 saccharomyc
113	6	0.7	58	1	YLC7_YEREN	P21212 yersinia en	186	0.7	153	1	ATP6_PSEPU	P25761 pseudomonas
114	6	0.7	61	1	YML2_ARCFU	O28071 archaeoglob	187	0.7	154	1	CTSR_BACSU	P37568 bacillus su
115	6	0.7	66	1	THIS_ECOLI	O32585 escherichia	188	0.7	154	1	Y406_CHLTR	O84411 chlamydia t
116	6	0.7	68	1	THIS_ECOLI	P17494 agkistrodon	189	0.7	154	1	Y116_NEIMB	O91429 neisseria m
117	6	0.7	69	1	YDH2_AGRKH	P77272 escherichia	190	0.7	154	1	Y645_NEIMA	O91429 neisseria m
118	6	0.7	70	1	YDH2_SALTY	O924t1 salmonella	191	0.7	154	1	Y116_NEIMB	O91429 neisseria m
119	6	0.7	71	1	ATP9_TRTIRU	P22151 neurospora	192	0.7	155	1	TPC2_DROME	P47949 drosophila
120	6	0.7	74	1	ATP9_TRTIRU	O01554 trichophyto	193	0.7	155	1	YIEI_DROME	P47949 drosophila
121	6	0.7	86	1	VEG_BACSV	P37466 bacillus su	194	0.7	155	1	YIEI_DROME	P47949 drosophila
122	6	0.7	93	1	IFHB_PASHA	P95519 pasteurella	195	0.7	156	1	BDOP_MOUSE	P33368 mus musculu
123	6	0.7	93	1	IFHB_PASHA	P75704 escherichia	196	0.7	156	1	BDOP_MOUSE	P33368 mus musculu
124	6	0.7	94	1	AVAI1_YEAST	P32450 saccharomyc	197	0.7	164	1	BDOP_MOUSE	P33368 mus musculu
125	6	0.7	94	1	VP10_NPVAC	P04870 autographa	198	0.7	165	1	MOAE_ANASP	P33286 rattus norv
126	6	0.7	95	1	S110_PIG	P04163 sus scrofa	199	0.7	165	1	OGT_MYCLE	P33286 rattus norv
127	6	0.7	96	1	FER_APHSA	P00250 aphanothece	200	0.7	165	1	OGT_MYCPA	P33286 rattus norv
128	6	0.7	96	1	S110_HUMAN	P08206 homo sapien	201	0.7	166	1	COFL_MYCTU	O8245 pyrobaculi
129	6	0.7	101	1	VE7_HPV21	P50779 human papil	202	0.7	166	1	COFL_HUMAN	O8245 pyrobaculi
130	6	0.7	102	1	TMOD_PSEME	O00459 pseudomonas	203	0.7	166	1	COFL_MOUSE	O8245 pyrobaculi
131	6	0.7	104	1	METU_ECOLI	P08338 escherichia	204	0.7	166	1	COFL_PIG	O8245 pyrobaculi
132	6	0.7	107	1	RL21_AERPE	P58077 aeropyrum p	205	0.7	166	1	CPXP_ECOLI	P45592 rattus norv
133	6	0.7	108	1	PTCA_BACST	O45402 bacillus st	206	0.7	166	1	DYR_STHHA	O54277 staphylococ
134	6	0.7	108	1	T2AG_ONCMY	O90y66 oncorhynch	207	0.7	166	1	RS5_BACSD	O92917 bacillus ha
135	6	0.7	110	1	Y103_ARATH	O80594 arabidopsis	208	0.7	166	1	RS5_BACST	O92917 bacillus ha
136	6	0.7	110	1	Y103_ARATH	O06783 puccinia gr	209	0.7	166	1	RS5_BACST	O92917 bacillus ha
137	6	0.7	113	1	USP1_PUCGR	P36227 puccinia gr	210	0.7	166	1	FTNB_ECOLI	P21467 bacillus su
138	6	0.7	113	1	Y544_METTH	O26644 methanobact	211	0.7	169	1	SCSB_HAETN	P32091 escherichia
139	6	0.7	113	1	Y544_METTH	O9122 cyanidium c	212	0.7	170	1	EIBS_ADEC2	P4853 haemophilus
140	6	0.7	115	1	YQ20_CAUCR	O92454 caulobacter	213	0.7	171	1	OGT_ECOLI	P35983 canine aden
141	6	0.7	115	1	YQ20_CAUCR	O78418 guillardia	214	0.7	171	1	OGT_ECOLI	P37429 salmonella
142	6	0.7	116	1	PSBW_GUTTH	P03339 felinus th	215	0.7	172	1	ESPA_LACVU	P35578 iacerta viv
143	6	0.7	117	1	WNT6_THUTH	O63396 rattus norv	216	0.7	173	1	ELHA_APLCA	P01360 aplysia cal
144	6	0.7	118	1	GAG_FSVGR	O92c90 rickettsia	217	0.7	174	1	HSCB_HAETN	O37006 haemophilus
145	6	0.7	119	1	P15_RAT	O22008 bacterioph	218	0.7	175	1	THIM_WHEAT	P20862 escherichia
146	6	0.7	119	1	RL22_RICPR	O22008 bacterioph	219	0.7	175	1	Y157_ARCFU	P45515 citrobacter
147	6	0.7	120	1	GTRA_BPSF2	O92442 bacterioph	220	0.7	176	1	Y157_ARCFU	O92442 bacterioph
148	6	0.7	120	1	GTRA_BPSF2	O92442 bacterioph	221	0.7	176	1	Y157_ARCFU	O92442 bacterioph
149	6	0.7	122	1	RL19_MYCA	O92442 bacterioph	222	0.7	177	1	Y157_ARCFU	O92442 bacterioph
150	6	0.7	122	1	RL7_BACST	O92442 bacterioph	223	0.7	178	1	Y157_ARCFU	O92442 bacterioph
151	6	0.7	124	1	MUNT_BUCAT	P05392 bacillus st	224	0.7	179	1	Y157_ARCFU	O92442 bacterioph
152	6	0.7	126	1	P15_HUMAN	P53999 homo sapien	225	0.7	180	1	EDGL_MOUSE	O52592 salmonella
153	6	0.7	126	1	P15_HUMAN	P11031 mus musculu	226	0.7	180	1	EDGL_MOUSE	O52592 salmonella
154	6	0.7	127	1	SY28_HUMAN	O9nrt3 homo sapien	227	0.7	181	1	FLAV_RHOCA	O93388 gallus gall
155	6	0.7	128	1	YDEP_BACSU	P96673 bacillus su	228	0.7	181	1	FLAV_RHOCA	O93388 gallus gall
156	6	0.7	130	1	H2A2_SCHPO	P04910 schizosacch	229	0.7	181	1	NOQ2_THETH	P52677 rhodobacter
157	6	0.7	131	1	ANFC_ANGJA	P18145 anguilla ja	230	0.7	181	1	NOQ2_THETH	P52677 rhodobacter
158	6	0.7	131	1	H2A1_SCHPO	P04909 schizosacch	231	0.7	181	1	NOQ2_THETH	P52677 rhodobacter
159	6	0.7	133	1	ANFC_SQVAC	P41319 squalus aca	232	0.7	182	1	YJH0_ECOLI	O39368 escherichia
160	6	0.7	135	1	RS9_PYPFU	O8u0e7 pyrococcus	233	0.7	182	1	HSB2_HUMAN	O16082 homo sapien
161	6	0.7	137	1	NIFB_KLEOX	P56265 klebsiella	234	0.7	182	1	HSB2_HUMAN	O16082 homo sapien
162	6	0.7	139	1	NIFB_KLEOX	P56265 klebsiella	235	0.7	182	1	HSB2_HUMAN	O16082 homo sapien
163	6	0.7	140	1	YEB6_MYCPN	P73381 mycoplasma	236	0.7	183	1	ATPD_BACHD	P01575 mus musculu
164	6	0.7	141	1	HBA_COLLI	P73381 mycoplasma	237	0.7	183	1	ATPD_BACHD	P01575 mus musculu
165	6	0.7	141	1	YAMA_RHISN	P21871 columba liv	238	0.7	186	1	DEF_CHLPM	O92612 chlamydia p
166	6	0.7	142	1	OSMC_ECOLI	P55560 rhizobium s	239	0.7	186	1	DEF_CHLPM	O92612 chlamydia p
167	6	0.7	142	1	OSMC_ECOLI	P55560 rhizobium s	240	0.7	187	1	YJH0_ECOLI	O39368 escherichia
168	6	0.7	145	1	Y187_METUA	O8245 pyrobaculum	241	0.7	187	1	YJH0_ECOLI	O39368 escherichia
169	6	0.7	145	1	Y187_METUA	O8245 pyrobaculum	242	0.7	187	1	YJH0_ECOLI	O39368 escherichia
170	6	0.7	146	1	LYCV_BPAPS	O91344 pseudomonas	243	0.7	187	1	YJH0_ECOLI	O39368 escherichia
171	6	0.7	146	1	LYCV_BPAPS	O91344 pseudomonas	244	0.7	187	1	YJH0_ECOLI	O39368 escherichia
172	6	0.7	147	1	GCSH_AERPE	O02260 saccharomyc	245	0.7	187	1	YJH0_ECOLI	O39368 escherichia
173	6	0.7	147	1	GCSH_AERPE	O02260 saccharomyc	246	0.7	188	1	YJH0_ECOLI	O39368 escherichia
174	6	0.7	147	1	PTVA_ECOLI	P33058 escherichia	247	0.7	188	1	YJH0_ECOLI	O39368 escherichia
175	6	0.7	148	1	Y110_MYCTU	O10827 mycobacteri	248	0.7	188	1	YJH0_ECOLI	O39368 escherichia
176	6	0.7	148	1	R12B_SCHPO	O74322 schizosacch	249	0.7	191	1	YJH0_ECOLI	O39368 escherichia
177	6	0.7	149	1	Y477_MYCTU	O11144 mycobacteri	250	0.7	191	1	YJH0_ECOLI	O39368 escherichia
178	6	0.7	149	1	CCMH_RHOCA	O00501 rhodobacter	251	0.7	191	1	YJH0_ECOLI	O39368 escherichia
179	6	0.7	149	1	MLE3_MOUSE	P05978 mus musculu	252	0.7	192	1	IL18_PIG	O19073 sus scrofa

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:24:59 ; Search time 16 seconds
(Without alignments)
2382.297 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 919
Sequence: 1 MSKPVLFANRSEMPVALAAY.....TGVKEEGNPILKHFITGTPP 919Scoring table: OLIGO
Gapop 60.0 , Capext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt-40:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	1.2	578	1 YTFM_HAETN	P44038 haemophilus
2	9	1.0	851	1 STR2_HUMAN	P52630 homo sapien
3	8	0.9	200	1 NAFQ_ECOLI	P31932 escherichia
4	8	0.9	360	1 KAG2_CAEBL	Q27535 caenorhabdi
5	8	0.9	577	1 YTFM_ECOLI	P39320 escherichia
6	8	0.9	734	1 MCCA_ARATH	Q42523 arabidopsis
7	8	0.9	1047	1 CARB_THENC	O9HK17 thermoplas
8	8	0.9	1355	1 SALM_DROME	P38770 drosophila
9	8	0.9	1402	1 SALM_DROME	P38806 drosophila
10	7	0.8	82	1 HYBG_ECOLI	P37185 escherichia
11	7	0.8	107	1 YGT2_YEAST	P53103 saccharomyc
12	7	0.8	109	1 Y353_MYCPN	P75249 mycoplasma
13	7	0.8	159	1 MP63_MYCPN	P97175 mycobacteri
14	7	0.8	190	1 GRA1_TOXCO	P14403 toxoplasma
15	7	0.8	190	1 RS7_AVIMR	O9ZNS1 avicennia m
16	7	0.8	195	1 INOI_BOVIN	P07352 bos taurus
17	7	0.8	211	1 FDOI_ECOLI	P32174 escherichia
18	7	0.8	211	1 FIXJ_AZOCA	P26487 azorhizobiu
19	7	0.8	223	1 VATD_THERM	O87880 thermus the
20	7	0.8	226	1 GPH_VIBCH	O9KNV6 vibrio chol
21	7	0.8	232	1 COX2_ASCSU	P24882 ascaris suu
22	7	0.8	242	1 RSTA_ECOLI	P56857 homo sapien
23	7	0.8	261	1 CUDI_MOUSE	P52106 escherichia
24	7	0.8	264	1 PURR_LACIA	O53065 lactococcus
25	7	0.8	271	1 LIGE_PSEPA	P27457 pseudomonas
26	7	0.8	280	1 LIGE_PSEPA	P39803 bacillus su
27	7	0.8	280	1 YTTT_BACSU	Q16762 homo sapien
28	7	0.8	296	1 TSTR_HUMAN	P52196 mus musculi
29	7	0.8	296	1 TSTR_MOUSE	P71556 mycobacteri
30	7	0.8	303	1 34KD_MYCTU	O96f65 homo sapien
31	7	0.8	313	1 OYE6_HUMAN	O8f6f5 mycobacteri
32	7	0.8	314	1 MRAM_FUSIN	O8f6f5 rhizobium m
33	7	0.8	317	1 MOCA_RHIME	P49307 thiazobium m

34	7	0.8	323	1 ISPB_ECOLI	P19641 escherichia
35	7	0.8	331	1 E13K_TOBAC	P52398 nicotiana t
36	7	0.8	332	1 HIT2_HAETN	P35755 haemophilus
37	7	0.8	347	1 YPEFG_ECOLI	P76559 escherichia
38	7	0.8	349	1 GP21_HUMAN	O99679 homo sapien
39	7	0.8	358	1 HRCA_CAUCR	P54305 caulobacter
40	7	0.8	375	1 PCE_TACMR	P21902 tachypleus
41	7	0.8	383	1 Y628_MYCTU	P66918 mycobacteri
42	7	0.8	386	1 ALR_ANASP	O8Y96 anabaena sp
43	7	0.8	397	1 WH42_STRCO	P23158 streptomyce
44	7	0.8	398	1 GDN_HUMAN	P07093 homo sapien
45	7	0.8	402	1 YDCM_ECOLI	P76102 escherichia
46	7	0.8	404	1 PRHA_METH	P19496 methanobact
47	7	0.8	404	1 DGT1_AGR75	O8uep3 agrobacteri
48	7	0.8	423	1 TTG_BACSU	P80698 bacillus su
49	7	0.8	428	1 C124_MYCTU	O50696 mycobacteri
50	7	0.8	433	1 YBBY_ECOLI	P77328 escherichia
51	7	0.8	454	1 YB48_MYCTU	O06548 mycobacteri
52	7	0.8	454	1 Y145_MYCTU	P95269 mycobacteri
53	7	0.8	481	1 INX6_DROME	O9YR82 drosophila
54	7	0.8	488	1 TPS1_KLUFA	Q07158 kluyveromyc
55	7	0.8	496	1 DID3_YEAST	P39976 saccharomyc
56	7	0.8	496	1 SRM_MOUSE	Q62270 mus musculi
57	7	0.8	498	1 FSC3_HUMAN	O9nqf6 homo sapien
58	7	0.8	507	1 MOSB_RHIME	O07608 rhizobium m
59	7	0.8	512	1 AMY_BACLI	P06278 bacillus li
60	7	0.8	513	1 YMT3_YEAST	Q04305 saccharomyc
61	7	0.8	519	1 AMYG_SACFI	P08017 saccharomyc
62	7	0.8	519	1 AMYH_SACFI	P26989 saccharomyc
63	7	0.8	522	1 MSAB_NEIMO	P14930 n peptide m
64	7	0.8	522	1 MSAB_NEIMA	O914m8 n peptide m
65	7	0.8	522	1 MSAB_NEIMB	O941m8 n peptide m
66	7	0.8	552	1 Y664_HAETN	O57538 haemophilus
67	7	0.8	567	1 IF37_SCHPO	O94236 schizosacch
68	7	0.8	571	1 YB63_SCHPO	Q09744 schizosacch
69	7	0.8	574	1 YHGE_ECOLI	P45804 escherichia
70	7	0.8	579	1 GPC2_RAT	P51653 rattus norv
71	7	0.8	587	1 DSD2_PSEAE	O91104 pseudomonas
72	7	0.8	595	1 REB1_KLUFA	O05950 kluyveromyc
73	7	0.8	613	1 ILVD_VIBCH	O9KWD0 vibrio chol
74	7	0.8	624	1 AMYG_ARXAD	P42042 arxula aden
75	7	0.8	628	1 DNLJ_THEMA	O9Kwv5 thermotoga
76	7	0.8	725	1 HS9B_BRARE	O57521 brachydanio
77	7	0.8	776	1 YOK5_CAEBL	O11177 caenorhabdi
78	7	0.8	807	1 MCM3_XENLA	P49739 xenopus lae
79	7	0.8	808	1 MCM3_HUMAN	P25206 homo sapien
80	7	0.8	812	1 YR86_CAEBL	O09568 caenorhabdi
81	7	0.8	817	1 HMT1_SCHPO	O02592 schizosacch
82	7	0.8	830	1 EXSB_CHLMO	O9p1t8 chlamydia m
83	7	0.8	1026	1 P11A_BOVIN	P32871 bos taurus
84	7	0.8	1068	1 P11A_HUMAN	P42336 homo sapien
85	7	0.8	1068	1 P11A_MOUSE	P42337 mus musculi
86	7	0.8	1068	1 YML7_YEAST	O03735 saccharomyc
87	7	0.8	1134	1 SLAP_BACST	P35825 bacillus st
88	7	0.8	1228	1 NRCA_CHICK	P35331 gallus gall
89	7	0.8	1284	1 FRPA_NEIMB	O9K0K9 neisseria m
90	7	0.8	1302	1 XDH_DROME	P10331 drosophila
91	7	0.8	1335	1 XDH_DROPS	P22811 drosophila
92	7	0.8	1342	1 XDH_DROSU	P91711 drosophila
93	7	0.8	1444	1 CA1E_HUMAN	P39059 homo sapien
94	7	0.8	1388	1 DP2L_PYRHO	O57661 pyrococcus
95	7	0.8	1431	1 RGSC_HUMAN	O14924 homo sapien
96	7	0.8	1447	1 CHD2_HUMAN	O14647 homo sapien
97	7	0.8	1739	1 FRPC_NEIMB	O914V5 neisseria m
98	7	0.8	1829	1 FRPC_NEIMC	P55127 neisseria m
99	7	0.8	1829	1 GP21_RAT	P16634 rattus norv
100	7	0.8	1866	1 BA72_HUMAN	P48634 homo sapien
101	7	0.8	2142	1 CCAB_RAT	O02284 rattus norv
102	7	0.8	2336	1 POLG_POWVL	Q04538 t genome po
103	7	0.8	3415	1 CY35_DESAC	P81079 desulfuromo
104	6	0.7	40	1 LA89_LACAC	O48501 lactobacilli
105	6	0.7	36	1 ATPB_SCHPO	P21536 schizosacch
106	6	0.7	48	1	

A48075
60K stress-related protein - chicken (Fragments)
C:Species: Gallus gallus (chicken)
C:Date: 16-Feb-1994 #sequence_revision 07-Feb-1997 #text_change 11-Jan-2000
C:Accession: A48075; B48075; D48075; E48075; F48075; G48075
R:Smith, D.F.; Sullivan, W.P.; Marlon, T.N.; Zaltsu, K.; Madden, B.; McCormick, D.J.; TC
Mol. Cell. Biol. 13, 869-876, 1993
A:Title: Identification of a 60-kilodalton stress-related protein, p60, which interacts
A:Reference number: A48075; MUID:93140785; PMID:8423808
A:Accession: A48075
A:Molecule type: protein
A:Residues: 1-36 <SM1>
A:Note: sequence extracted from NCBI backbone (NCBIP:123364)
A:Accession: B48075
A:Molecule type: protein
A:Residues: 37-43 <SM2>
A:Note: sequence extracted from NCBI backbone (NCBIP:123366)
A:Accession: C48075
A:Molecule type: protein
A:Residues: 44-59 <SM3>
A:Note: sequence extracted from NCBI backbone (NCBIP:123368)
A:Accession: D48075
A:Molecule type: protein
A:Residues: 60-73 <SM4>
A:Note: sequence extracted from NCBI backbone (NCBIP:123370)
A:Accession: E48075
A:Molecule type: protein
A:Residues: 74-98 <SM5>
A:Note: sequence extracted from NCBI backbone (NCBIP:123372)
A:Accession: F48075
A:Molecule type: protein
A:Residues: 99-108 <SM6>
A:Note: sequence extracted from NCBI backbone (NCBIP:123374)
A:Accession: G48075
A:Molecule type: protein
A:Residues: 109-124 <SM7>
A:Note: sequence extracted from NCBI backbone (NCBIP:123376)
C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat hc

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 124;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 387 RELLEQL 393
|||||||
DB 52 RELLEQL 58

RESULT 38
F97593
50S ribosomal protein L22 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: F97593
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: F97593
A:Molecule type: DNA
A:Residues: 1-129 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87703.1; PID:915157064; GSPDB:GN00169
A:Gene: AGR_C_3548
A:Map position: circular chromosome
C:Superfamily: Escherichia coli ribosomal protein L22

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 129;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 491 KUNLVAA 497
|||||||
DB 28 KUNLVAA 34

RESULT 39
AC2815
50S ribosomal protein L22 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AC2815
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; MCCI
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AC2815
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-129 <KUR>
A:Cross-references: GB:AE00688; PIDN:ALA2937.1; PID:917740394; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: rplV
A:Map position: circular chromosome
C:Superfamily: Escherichia coli ribosomal protein L22

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 129;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 491 KUNLVAA 497
|||||||
DB 28 KUNLVAA 34

RESULT 40
AC1545
hypothetical protein lin0899 [imported] - Listeria innocua (strain C1p11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AC1545
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitouram, A.;
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1545
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-130 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC96131.1; PID:916413349; GSPDB:GN00178
A:Experimental source: strain C1p11262
C:Genetics:
A:Gene: lin0899

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 130;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 698 TOAPPET 704
|||||||
DB 15 TOAPPET 21

Search completed: April 28, 2003, 16:36:03
Job time : 78 secs

submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T18004
A:Status: preliminary; translated from GB/EMBL/DDAJ

A:Molecule type: DNA
A:Residues: 1-95 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96869.1

A:Experimental source: specific host *Chloroella* strain NC64A
C:Genetics:
A:Note: A502L

Query Match 0.8%; Score 7; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 817 QSRGYA 823
|||||
Db 47 QSRGYA 53

RESULT 33
S61135
hypothetical protein YGL182c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein G1607
C:Species: *Saccharomyces cerevisiae*
C:Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 19-Apr-2002

C:Accession: S61135; S64199
R:Bertani, I.; Cogillevina, M.; Zaccaria, P.; Klima, R.; Brusch, C.V.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch

A:Reference number: A69441
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-119 <KLE>
A:Cross-references: GB:AE000997; GB:AE000782; NID:g2689320; PIDN:AAB89718.1; PID:g264

A:Experimental source: strain S288C
C:Superfamily: conserved hypothetical protein MJ0039

Query Match 0.8%; Score 7; DB 1; Length 119;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 LRPEQIQ 67
|||||
Db 102 LRPEQIQ 108

RESULT 36
A97825
50S ribosomal protein L22 (imported) - *Rickettsia conorii* (strain Malish 7)

C:Species: *Rickettsia conorii*
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: A97825
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fourmiller, P.E.; Barbe, V.; Samson, D.;
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in *Rickettsia conorii* and *Rickettsia prowazekii*.

A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: A97825
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-119 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL03539.1; PID:g15620116; GSPDB:GN00173

C:Genetics:
A:Gene: rplV
C:Superfamily: *Escherichia coli* ribosomal protein L22

Query Match 0.8%; Score 7; DB 2; Length 119;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 491 KLNLYAA 497
|||||
Db 22 KLNLYAA 28

RESULT 37
S73639
MG553 homolog g12_orf109 - *Mycoplasma pneumoniae* (strain ATCC 29342)

C:Species: *Mycoplasma pneumoniae*
A:Variety: ATCC 29342
C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999

C:Accession: S73639
R:Himmelfeich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*
A:Reference number: S73327; MUID:97105885; PMID:8948633
A:Accession: S73639
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-109 <HIM>

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AC00884
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-82 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD02975.1; PID:g16504223; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY3314
 C:Superfamily: hydrogenase expression/formation protein hycp

Query Match 0.8%; Score 7; DB 2; Length 82;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 839 GOVLAVG 845
 Db 7 GOVLAVG 13

RESULT 28
 C85958
 hypothetical protein hycG [imported] - *Escherichia coli* (strain O157:H7, substrain EDL93
 C:Species: *Escherichia coli*
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
 C:Accession: C85958
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 11ler, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: AB5480; MUID:21074935; PMID:11206551
 A:Accession: C85958
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 182 <STO>
 A:Cross-references: GB:AE005174; MID:g12517554; PIDN:AAG58127.1; GSPDB:GN00145; UWGP:243
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: hycG
 C:Superfamily: hydrogenase expression/formation protein hycp

Query Match 0.8%; Score 7; DB 2; Length 82;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 839 GOVLAVG 845
 Db 7 GOVLAVG 13

RESULT 29
 C91113
 hydrogenase-2 operon protein hycG [imported] - *Escherichia coli* (strain O157:H7, substr
 C:Species: *Escherichia coli*
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 17-May-2002
 C:Accession: C91113
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gawara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genc
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: C91113
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-82 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA37298.1; PID:g13363347; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 050952
 C:Genetics:
 A:Gene: ECS3875
 C:Superfamily: hydrogenase expression/formation protein hycp

Query Match 0.8%; Score 7; DB 2; Length 82;

Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 839 GOVLAVG 845
 Db 7 GOVLAVG 13

RESULT 30
 AH0088
 probable flagellar regulatory protein YPO0720 [imported] - *Yersinia pestis* (strain CO
 C:Species: *Yersinia pestis*
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
 C:Accession: AH0088
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AH0088
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-88 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC89571.1; PID:g15978804; GSPDB:GN00175
 C:Genetics:
 A:Gene: YPO0720

Query Match 0.8%; Score 7; DB 2; Length 88;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 533 AVARAIL 539
 Db 77 AVARAIL 83

RESULT 31
 AB1784
 hypothetical protein NMA2123 [imported] - *Neisseria meningitidis* (strain Z2491 serogr
 C:Species: *Neisseria meningitidis*
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: AB1784
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
 ; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
 Nature 404, 502-506, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491
 A:Reference number: AB1775; MUID:20222556; PMID:10761919
 A:Accession: AB1784
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-91 <PAR>
 A:Cross-references: GB:AL162758; GB:AL157959; MID:g7380672; PIDN:CAB85336.1; PID:g738
 A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: NMA2123

Query Match 0.8%; Score 7; DB 2; Length 91;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 LRPEQIO 67
 Db 8 LRPEQIO 14

RESULT 32
 T18004
 hypothetical protein A502L - *Chlorella virus PBCV-1*
 C:Species: *Chlorella virus PBCV-1*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18004
 R:Graves, M.V.; Van Etten, J.L.

C:Accession: T00661
 R:Pederspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo, R.; Vysotskaya, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.
 Submitted to the EMBL Data Library, February 1998
 A:Reference number: Z14197
 A:Accession: T00661
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1417 <PED>
 A:Cross-references: EMBL:AC002396; NID:g2749918; PID:g2829883; GSPDB:GN00059; ATSP:F316.
 C:Genetics:
 A:Gene: ATSP:F316.24
 A:Map position: 1
 A:Introns: 23/1; 46/3; 276/1; 309/1; 546/1; 1269/1; 1284/2; 1333/1
 Query Match 0.9%; Score 8; DB 2; Length 1417;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 387 RELLEQL 394
 |||||
 Db 843 RELLEQL 850

RESULT 24

hypothetical protein XF1543 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: G82667
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below

A:Accession: G82667
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-66 <SIM>

A:Cross-references: GB:AE003984; GB:AE003849; NID:g9106581; PIDN:AAE4352.1; GSPDB:GN001

A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Brlones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carriero, D.M.; Carrer, H

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincan, A.P.; Ferreira, A.J.S.
 Submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tsunako, M.H.; Vallada, H.; Van Sluys, W.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328
 A:Contents: annotation

C:Genetics:
 A:Gene: XF1543

Query Match 0.8%; Score 7; DB 2; Length 66;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 809 LRFAGG 815
 |||||
 Db 11 LRFAGG 17

RESULT 25

hypothetical protein ssc7201 - Synechocystis sp. (strain PCC 6803)
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S75519
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O.K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys

S.
 A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S75519
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-72 <KAN>

A:Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BA18080.1; PID:dl01

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 0.8%; Score 7; DB 2; Length 72;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 216 LBDITQE 222
 |||||
 Db 54 LBDITQE 60

RESULT 26

hydrogenase-2 operon protein hylg - Escherichia coli (strain K-12)

C:Species: Escherichia coli
 C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 01-Mar-2002

C:Accession: G55516; D65085
 R:Memon, N.K.; Chetels, C.Y.; Dervartanian, M.; Wendt, J.C.; Shannugam, K.T.; Peck J

J. Bacteriol. 176, 4416-4423, 1994
 A:Title: Cloning, sequencing, and mutational analysis of the hyl operon encoding Esch

A:Reference number: A55516; MUID:94292472; PMID:8021226
 A:Accession: G55516

A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-82 <MEN>

A:Cross-references: GB:U09177; NID:g501051; PIDN:AAA21595.1; PID:g544489
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: D65085
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1-82 <BLAT>

A:Cross-references: GB:AE000382; GB:U00096; NID:g2367182; PIDN:AACT6026.1; PID:g17693

A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:

A:Gene: hylg
 C:Superfamily: hydrogenase expression/formation protein hylc

Query Match 0.8%; Score 7; DB 2; Length 82;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 839 GOVLAVG 845
 |||||
 Db 7 GOVLAVG 13

hydrogenase-2 component protein [imported] - Salmonella enterica subsp. enterica sero

C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 17-May-2002
 C:Accession: AG0884

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001

A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: F86119
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-577 <STO>
A:Cross-references: GB:AE005174; NID:g12519219; PIDN:AGS9418.1; GSPDB:GN00145; UMG:258
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ytfm

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 577;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 809 LRFPGAD 816
|||||||
DB 466 LRFPGAD 473

RESULT 19
F82064

conserved hypothetical protein VC2548 [imported] - *Vibrio cholerae* (strain N16961 serogr
C:Species: *Vibrio cholerae*
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82064
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
charlson, D.; Ermolaeva, M.D.; Vamthyan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406: 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: F82064
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-582 <HEI>
A:Cross-references: GB:AE004323; GB:AE003852; NID:g9657119; PIDN:AF95689.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2548
A:Map position: 1

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 582;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 809 LRFPGAD 816
|||||||
DB 473 LRFPGAD 480

RESULT 20
AH2546

hypothetical protein all7659 [imported] - *Nostoc* sp. (strain PCC 7120) plasmid PCC7120p6
C:Species: *Nostoc* sp.
A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AH2546
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuch
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the filamentous Nitrogen-fixing Cyanobacterium *Ana*
A:Reference number: AB1807; MUID:21595285; PMID:11755840
A:Accession: AH2546
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-731 <KUR>
A:Cross-references: GB:AP003602; PIDN:BAW77302.1; PID:g17134744; GSPDB:GN00181
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all7659
A:Genome: plasmid

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 731;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 IKALEDI 219
|||||||
DB 116 IKALEDI 123

RESULT 21
S40022

spalt protein - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C>Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000
C:Accession: S40022
R:Kuehnlein, R.P.; Frommer, G.; Friedrich, M.; Gonzalez-Gaitan, M.; Weber, A.; Wagner
EMBO J. 13, 168-179, 1994
A:Title: spalt encodes an evolutionarily conserved zinc finger protein of novel struc
A:Reference number: S40022; MUID:94139659; PMID:7905822
A:Accession: S40022
A:Molecule type: DNA
A:Residues: 1-1355 <KUE>
A:Cross-references: EMBL:X75541; NID:g414106; PIDN:CA53229.1; PID:g2598394
A:Note: the authors translated the codon GAC for residue 51 as Ala
A:Gene: sal; spalt
A:Cross-references: FlyBase:FBgn0004579
A:Introns: 51/1; 1329/3 1355/2
C:Keywords: zinc finger

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 1355;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LTFEIOA 68
|||||||
DB 913 LTFEIOA 920

RESULT 22
S42748

finger protein - fruit fly (*Drosophila virilis*) (fragment)
C:Species: *Drosophila virilis*
C>Date: 20-Oct-1994 #sequence_revision 26-May-1995 #text_change 24-Sep-1998
C:Accession: S42748
R:Schuh, R.
submitted to the EMBL Data Library, November 1993
A:Reference number: S42748
A:Accession: S42748
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1402 <SCH>
A:Cross-references: EMBL:227444; NID:g426461; PID:g426462
C:Genetics:
A:Gene: FlyBase:FlyBase:FBgn0013137
A:Cross-references: FlyBase:FBgn0013137
C:Keywords: DNA binding; transcription regulation

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 1402;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LTFEIOA 68
|||||||
DB 934 LTFEIOA 941

RESULT 23
T00661

hypothetical protein F316.24 - *Arabidopsis thaliana*
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 22-Oct-1999

Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 492 LNLVAKA 499
|||||||
Db 254 LNLVAKA 261

RESULT 14

JC5747
coronafacic acid synthetase component cfa3 [imported] - Pseudomonas syringae

C:Contains: 3-oxoacyl-[acyl-carrier-protein] synthase (EC 3.2.1.-)

C:Species: Pseudomonas syringae

C>Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000

C:Accession: JC5747

R:Penfold, C.N.; Bender, C.L.; Turner, J.G.

Gene 183, 167-173, 1996

A:Title: Characterisation of genes involved in biosynthesis of coronafacic acid, the pol

A:Reference number: JC5745; MUID:97149295; PMID:8996103

A:Accession: JC5747

A:Molecule type: DNA

A:Residues: 1-380 <PEN>

A:Cross-references: GB:U56980; MID:91655810; PIDN:AA841300.1; PID:91655813

A>Note: the authors translated the initiation codon GTG for residue 1 as Met

C:Genetics:

A:Gene: cfa3

A:Start codon: GTG

C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I; 3-oxoacyl-[acyl-carrier-prot

C:Keywords: glycosidase; hydrolase; transferase

F:26-372/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>

Query Match 0.9%; Score 8; DB 2; Length 380;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 529 DAYSAAVAR 536
|||||||
Db 230 DAYSAAVAR 237

RESULT 15

T34645
hypothetical protein SC10H5.07 SC10H5.07 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T34645

R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, August 1998

A:Reference number: 221550

A:Accession: T34645

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-469 <COLI>

A:Cross-references: EMBL:AL031232; PIDN:CAA20279.1; GSPDB:GN00070; SCODEB:SC10H5.07

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCODEB:SC10H5.07

Query Match 0.9%; Score 8; DB 2; Length 469;

Best Local Similarity 100.0%; Pred. No. 24;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 PVLAAYL 21
|||||||

Db 420 PVLAAYL 427

RESULT 16

S56445

hypothetical 64.8k protein (msra-chpbl intergenic region) - Escherichia coli (strain K-1

N:Alternate names: hypothetical protein o577

C:Species: Escherichia coli

C>Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 01-Mar-2002

C:Accession: S56445; G65233

R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.

Nucleic Acids Res. 23, 2105-2119, 1995

A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from

A:Reference number: S56314; MUID:95334362; PMID:7610040

A:Accession: S56445

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-577 <BUR>

A:Cross-references: EMBL:U14003; MID:91263172; PIDN:AAA97116.1; PID:9537061

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G65233

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-577 <BLAT>

A:Cross-references: GB:AE000493; GB:U00096; MID:92367360; PIDN:AACT7177.1; PID:917906

A:Experimental source: strain K-12, substrain MGI655

C:Genetics:

A:Gene: ytfM

A:Start codon: GTG

Query Match 0.9%; Score 8; DB 2; Length 577;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 809 LRFFAGCD 816
|||||||
Db 466 LRFFAGCD 473

RESULT 17

F91278
hypothetical protein ECs5198 [imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: F91278

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9

A:Reference number: A9629; MUID:21156231; PMID:11258796

A:Accession: F91278

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-577 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA838621.1; PID:913364675; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs5198

Query Match 0.9%; Score 8; DB 2; Length 577;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 809 LRFFAGCD 816
|||||||

Db 466 LRFFAGCD 473

RESULT 18

F86119

hypothetical protein ytfM [imported] - Escherichia coli (strain O157:H7, substrain ED

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: F86119

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

RESULT 9
E85859
cytochrome c-type protein [imported] - Escherichia coli (strain O157:H7, substrain EDL93)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85859
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: E85859
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-200 <STO>
A:Cross-references: GB:AE005174; NID:g12516534; PIDN:AAG57337.1; GSPDB:GN00145; UWGP:234
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: napC
C:Superfamily: denitrification system component nirt; nirt homology

Query Match
Best Local Similarity 0.9%; Score 8; DB 2; Length 200;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 161 KPGIKRL 168
Db 7 KPGIKRL 14

RESULT 10
AE1390
MreB-like protein homolog mbl [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AE1390
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AE1390
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-331 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAD00603.1; PID:g16412013; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: mbl
C:Superfamily: rod shape-determining protein envB

Query Match
Best Local Similarity 0.9%; Score 8; DB 2; Length 331;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 839 GOVLAVGT 846
Db 38 GOVLAVGT 45

RESULT 11
AG1765
MreB-like protein homolog mbl [imported] - Listeria innocua (strain Cl1p11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AG1765
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AG1765
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-331 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC97895.1; PID:g16415205; GSPDB:GN00178
A:Experimental source: strain Cl1p11262
C:Genetics:
A:Gene: mbl
C:Superfamily: rod shape-determining protein envB

Query Match
Best Local Similarity 0.9%; Score 8; DB 2; Length 331;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 839 GOVLAVGT 846
Db 38 GOVLAVGT 45

Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AG1765
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-331 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC97895.1; PID:g16415205; GSPDB:GN00178
A:Experimental source: strain Cl1p11262
C:Genetics:
A:Gene: mbl
C:Superfamily: rod shape-determining protein envB

Query Match
Best Local Similarity 0.9%; Score 8; DB 2; Length 331;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 839 GOVLAVGT 846
Db 38 GOVLAVGT 45

RESULT 12
T44739
hypoetical protein MLCB1450.24 [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T44739
R:James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
Submitted to the EMBL data library, January 1998
A:Reference number: Z22831
A:Accession: T44739
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-340 <JAM>
A:Cross-references: EMBL:AL035159; PIDN:CAA22708.1
A:Experimental source: cosmid B1450
C:Genetics:
A:Note: MLCB1450.24

Query Match
Best Local Similarity 0.9%; Score 8; DB 2; Length 340;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 884 GVRWSPV 891
Db 167 GVRWSPV 174

RESULT 13
T27569
hypoetical protein ZC434.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C:Accession: T27569
R:Milkinson, J.
Submitted to the EMBL data library, July 1996
A:Reference number: Z20388
A:Accession: T27569
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-360 <WIL>
A:Cross-references: EMBL:Z75714; PIDN:CAB00062.1; GSPDB:GN00019; CESP:ZC434.8
A:Experimental source: clone ZC434
C:Genetics:
A:Gene: CESP:ZC434.8
A:Map position: 1
A:Introns: 13/3; 109/3; 310/3
C:Superfamily: creatine kinase; creatine kinase repeat homology

Query Match
Best Local Similarity 0.9%; Score 8; DB 2; Length 360;

A46160
interferon alpha-induced transcription activator ISGF-3, 113K chain - human
N:Alternate names: stat2 protein
C:Species: Homo sapiens (man)
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Aug-2000
C:Accession: A46160; S71908; S53873
R:Fu, X.Y.; Schindler, C.; Improtta, T.; Abersold, R.; Darnell Jr., J.E.
Proc. Natl. Acad. Sci. U.S.A. 89, 7840-7843, 1992
A:Title: The proteins of ISGF-3, the interferon alpha-induced transcriptional activator,
A:Reference number: A46160; MUID:92366558; PMID:1502204
A:Accession: A46160
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA, protein
A:Residues: 1-851 <F01>
A>Note: sequence extracted from NCBI backbone (NCBIP:110820)
R:Yan, R.; Qureshi, S.; Zhong, Z.; Wen, Z.; Darnell, J.E.
Submitted to the EMBL data library, December 1994
A:Reference number: S71908
A:Accession: S71908
A:Molecule type: DNA
A:Residues: 1-851 <YAN>
A:Cross-references: EMBL:U18671; NID:91293919; PIDN:AAA98760.1; PID:91293920
R:Yan, R.; Qureshi, S.; Zhong, Z.; Wen, Z.; Darnell Jr., J.E.
Nucleic Acids Res. 23, 459-463, 1995
A:Title: The genomic structure of the STAT genes: multiple exons in coincident sites in
A:Reference number: S53873; MUID:95192056; PMID:7885841
A:Accession: S53873
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-196;392-591;684-730 <YAN>
A:Cross-references: EMBL:U18671
C:Genetics:
A:Gene: stat2
A:Introns: 44/2; 95/3; 127/3; 157/3; 183/1; 211/3; 261/2; 314/2; 345/2; 365/2; 372/2; 40
A:Superfamily: human signal transducer and transcription activator STAT3A
C:Keywords: signal transduction; transcription regulation

Query Match 1.0%; Score 9; DB 2; Length 851;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ESTLEPVE 465
|||||
DB 756 ESTLEPVE 764

RESULT 6
A:0651
hypothetical protein RV3072c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
R:COLE, S.T.; Broesh, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felkwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70500
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-174 <CO>
A:Cross-references: GB:28386; GB:AL123456; NID:93261691; PIDN:CAB06251.1; PID:e290922;
C:Genetics:
A:Gene: RV3072c

Query Match 0.9%; Score 8; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 895 RVDVATGV 902

DB 90 RVDVATGV 97
|||||

RESULT 7
H64989
membrane-bound tetraheme cytochrome napC - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A:Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H64989
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-200 <BLAT>
A:Cross-references: GB:A000309; GB:U00096; NID:91788520; PIDN:AAC75262.1; PID:917885
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: napC
C:Superfamily: denitrification system component nirT; nirT homology
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein
F:18-192/Domain: nirT homology <NIR>
F:57.60/Binding site: heme (Cys) (covalent) #status predicted
F:61/Binding site: heme iron (His) (axial ligand) #status predicted
F:87.90/Binding site: heme (Cys) (covalent) #status predicted
F:91/Binding site: heme iron (His) (axial ligand) #status predicted
F:147.150/Binding site: heme (Cys) (covalent) #status predicted
F:151/Binding site: heme iron (His) (axial ligand) #status predicted
F:179.182/Binding site: heme (Cys) (covalent) #status predicted
F:183/Binding site: heme iron (His) (axial ligand) #status predicted

Query Match 0.9%; Score 8; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 161 KPGLIKRL 168
|||||
DB 7 KPGLIKRL 14

RESULT 8
C91015
cytochrome c-type protein [imported] - Escherichia coli (strain O157:H7, substrain RI
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: C91015
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gatawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C91015
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-200 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA836514.1; PID:q13362560; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: Ecs3091
C:Superfamily: denitrification system component nirT; nirT homology

Query Match 0.9%; Score 8; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 161 KPGLIKRL 168
|||||
DB 7 KPGLIKRL 14

```
979 6 0.7 247 2 T14818 leucine-rich repea
980 6 0.7 247 2 T23763 hypothetical prote
981 6 0.7 247 2 AC3412 DNA repair protei
982 6 0.7 248 1 TWMVRS transforming prote
983 6 0.7 248 2 G83253 probable short-cha
984 6 0.7 248 2 F64938 hypothetical prote
985 6 0.7 248 2 C90940 hypothetical prote
986 6 0.7 248 2 G85788 hypothetical prote
987 6 0.7 248 2 AC0712 hypothetical prote
988 6 0.7 248 2 AC0712 probable outer mem
989 6 0.7 248 2 S76774 thiol-disulfide in
990 6 0.7 249 2 AD0394 hypothetical prote
991 6 0.7 249 2 B83767 glucose 1-dehydrog
992 6 0.7 249 2 J00274 conserved hypothet
993 6 0.7 249 2 AF3225 hypothetical 29k p
994 6 0.7 250 1 OBNC2 hypothetical prote
995 6 0.7 250 2 AB0171 cytochrome-c oxida
996 6 0.7 250 2 E84321 probable 3-deoxy-m
997 6 0.7 250 2 H84212 50S ribosomal prot
998 6 0.7 250 2 C95111 hypothetical prote
999 6 0.7 250 2 F75219 hypothetical prote
1000 6 0.7 250 2 B40626 probable 2,3-dihyd
```

ALIGNMENTS

RESULT 1

AE0428
Probable exported protein YPO3524 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AE0428

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tlthall, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Terraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 533-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AE0428

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1578 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC92753.1; PID:g15981448; GSPDB:GN00175

C:Genetics:

A:Gene: YPO3524

```
Query Match      1.4%: Score 13; DB 2; Length 578;
Best Local Similarity 100.0%; Pred. No. 0.00021;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 880 GAGVGRMASPVG 892

Db 538 GAGVGRMASPVG 550

RESULT 2

AD1055

Probable exported protein ytfm [imported] - Salmonella enterica subsp. enterica serovar

C:Species: Salmonella enterica subsp. enterica serovar typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001

C:Accession: AD1055

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Garra, P.

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; PMID:11677608

A:Accession: AD1055

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1577 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD06889.1; PID:g16505537; GSPDB:GN00176

C:Genetics:

A:Gene: ytfm

```
Query Match      1.2%: Score 11; DB 2; Length 577;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 882 GGVGRMASPVG 892

Db 539 GGVGRMASPVG 549

RESULT 3

B64012
hypothetical protein H10698 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999

C:Accession: B64012

R:Flitschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: B64012

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Cross-references: GB:U32752; GB:L42023; MID:g1573692; PIDN:AAC22357.1; PID:g1573700

```
Query Match      1.2%: Score 11; DB 2; Length 578;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 882 GGVGRMASPVG 892

Db 539 GGVGRMASPVG 549

RESULT 4

F83327

Conserved hypothetical protein PA2543 [imported] - Pseudomonas aeruginosa (strain PAO

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: F83327

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laribig, K.; L

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: F83327

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1579 <STO>

A:Cross-references: GB:AE004682; GB:AE004091; MID:g9948598; PIDN:AAG05931.1; GSPDB:GN

C:Experimental source: strain PA01

C:Genetics:

A:Gene: PA2543

```
Query Match      1.1%: Score 10; DB 2; Length 579;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 809 LRRFAGGDOS 818

Db 470 LRRFAGGDOS 479

RESULT 5

833	6	0.7	218	2	H64638	conserved hypotet
834	6	0.7	218	2	AH2149	hypothetical prote
835	6	0.7	218	2	B97428	hypothetical prote
836	6	0.7	218	2	G82019	probable periplasm
837	6	0.7	219	2	C75031	lsu ribosomal prot
838	6	0.7	219	2	D71217	probable ribosomal
839	6	0.7	219	2	B98348	probable transcrip
840	6	0.7	219	2	H81824	hypothetical perip
841	6	0.7	219	2	A82498	CbdY family protei
842	6	0.7	219	2	AE0945	probable exported
843	6	0.7	220	1	QOVZC9	F9 protein - sheep
844	6	0.7	220	1	S48551	hypothetical prote
845	6	0.7	220	2	T80329	receptor tyrosine
846	6	0.7	220	2	B72715	hypothetical prote
847	6	0.7	220	2	C72716	hypothetical prote
848	6	0.7	221	2	H83488	probable permease
849	6	0.7	221	2	D91210	probable phosphata
850	6	0.7	221	2	G86056	hypothetical prote
851	6	0.7	221	2	D65174	hypothetical prote
852	6	0.7	221	2	AG0956	probable hydrolase
853	6	0.7	221	2	D90195	hypothetical prote
854	6	0.7	221	2	T52622	probable peptidylp
855	6	0.7	222	1	B69394	probable phosphoe
856	6	0.7	222	2	AH0498	probable haloacid
857	6	0.7	222	2	D75549	sana protein - Dei
858	6	0.7	222	2	H87496	hypothetical prote
859	6	0.7	222	2	E70335	hypothetical prote
860	6	0.7	222	2	S77557	hypothetical prote
861	6	0.7	222	2	B97118	hypothetical prote
862	6	0.7	223	2	S39883	endo-1,4-beta-xyla
863	6	0.7	223	2	H75600	hypothetical prote
864	6	0.7	224	2	D70665	probable urease ac
865	6	0.7	224	2	S65960	maum protein precu
866	6	0.7	224	2	AF3382	alpha/beta hydroa
867	6	0.7	224	2	A64418	hypothetical prote
868	6	0.7	224	2	C91083	hypothetical membr
869	6	0.7	224	2	B75299	sensory transducl
870	6	0.7	224	2	T10660	photosystem II pro
871	6	0.7	224	2	T51875	hypothetical prote
872	6	0.7	225	2	T49769	hypothetical prote
873	6	0.7	225	2	AC1630	cell-division inh
874	6	0.7	225	2	A11267	cell-division inh
875	6	0.7	226	2	T11368	H+-transporting tw
876	6	0.7	226	2	S04752	H+-transporting tw
877	6	0.7	226	2	AF0665	probable amino aci
878	6	0.7	226	2	FA5239	cell-division inh
879	6	0.7	227	2	H75506	oxidoreductase, sh
880	6	0.7	227	2	AG0364	probable two compo
881	6	0.7	227	2	S44836	KO2D10.4 protein -
882	6	0.7	227	2	H90527	glucose inhibited
883	6	0.7	227	2	C83046	probable transcrip
884	6	0.7	228	2	AF0374	probable ABC trans
885	6	0.7	228	2	AF0253	cell division inh
886	6	0.7	228	2	D71359	conserved hypotet
887	6	0.7	228	2	T12799	conserved hypotet
888	6	0.7	228	2	B69777	hypothetical prote
889	6	0.7	228	2	AC2297	hypothetical prote
890	6	0.7	228	2	D90139	conserved hypotet
891	6	0.7	228	2	T48027	hypothetical prote
892	6	0.7	228	2	T45311	hypothetical prote
893	6	0.7	229	2	F70800	cell division alp-
894	6	0.7	229	2	H72318	probable transport
895	6	0.7	229	2	T20848	hypothetical prote
896	6	0.7	229	2	T48636	hypothetical prote
897	6	0.7	229	2	T48636	hypothetical prote
898	6	0.7	230	2	C75301	rRNA methylase Spo
899	6	0.7	230	2	B81397	probable 1-acylgly
900	6	0.7	230	2	D81252	hypothetical prote
901	6	0.7	232	2	D84028	septom site-determ
902	6	0.7	232	2	G84382	cobalt transport A
903	6	0.7	232	2	G83609	probable transcrip
904	6	0.7	232	2	E75547	hypothetical prote
905	6	0.7	232	2	E85928	hypothetical prote
906	6	0.7	233	2	T48179	hypothetical prote
907	6	0.7	233	2	H86925	probable lipoprote
908	6	0.7	233	2	E96903	distantly related
909	6	0.7	234	1	G69024	pheromone shutdown
910	6	0.7	234	2	PC4398	glutamine-tRNA lig
911	6	0.7	234	2	T06995	probable MADS box
912	6	0.7	234	2	T44731	hypothetical prote
913	6	0.7	234	2	F69019	conserved hypotet
914	6	0.7	234	2	A70736	hypothetical prote
915	6	0.7	234	2	D83403	hypothetical prote
916	6	0.7	234	2	A86290	tlfN11.11 protein
917	6	0.7	234	2	C75368	conserved hypotet
918	6	0.7	234	2	S60885	ferric exochelin u
919	6	0.7	234	2	AH3335	probable membrane-
920	6	0.7	235	2	A72639	probable uracil ph
921	6	0.7	235	2	C83057	conserved hypotet
922	6	0.7	235	2	B85639	hypothetical prote
923	6	0.7	235	2	A86150	tlfN6.24 protein -
924	6	0.7	235	2	AH3210	conserved hypotet
925	6	0.7	236	1	C27058	carboxymethylenebu
926	6	0.7	236	2	S02022	carboxymethylenebu
927	6	0.7	236	2	T44622	carboxymethylenebu
928	6	0.7	236	2	T44669	carboxymethylenebu
929	6	0.7	236	2	S60390	probable membrane
930	6	0.7	236	2	F75375	nodulin 21-related
931	6	0.7	236	2	B86881	hypothetical prote
932	6	0.7	236	2	T50908	hypothetical prote
933	6	0.7	237	2	D71877	probable amino aci
934	6	0.7	237	2	C64637	amino acid ABC tra
935	6	0.7	237	2	T07820	hypothetical prote
936	6	0.7	237	2	AE1757	phage related prot
937	6	0.7	237	2	AE1757	phage protein homo
938	6	0.7	237	2	H83033	probable transcrip
939	6	0.7	237	2	S75270	hypothetical prote
940	6	0.7	238	2	F87605	hypothetical prote
941	6	0.7	238	2	B84951	hypothetical prote
942	6	0.7	238	2	D96026	probable branched-
943	6	0.7	238	2	AH3176	conserved hypotet
944	6	0.7	239	2	T23169	conserved hypotet
945	6	0.7	239	2	B87062	transcription regu
946	6	0.7	239	2	A75585	probable membrane
947	6	0.7	239	2	AE0784	hypothetical prote
948	6	0.7	240	2	G73295	probable ATP-bind
949	6	0.7	240	2	F83488	probable amino aci
950	6	0.7	240	2	H95873	hypothetical prote
951	6	0.7	240	2	H97783	probable ATP-bind
952	6	0.7	240	2	G98155	hypothetical prote
953	6	0.7	240	2	AE3132	deoxynucleotide mo
954	6	0.7	241	1	KIBPDA	ribose-phosphate
955	6	0.7	241	2	C47019	probable octopine
956	6	0.7	241	2	F42600	hypothetical prote
957	6	0.7	241	2	AG3230	exopolysaccharide
958	6	0.7	241	2	A12588	conserved hypotet
959	6	0.7	241	2	F69150	conserved hypotet
960	6	0.7	241	2	AC2646	hypothetical prote
961	6	0.7	242	2	T13190	hypothetical prote
962	6	0.7	242	2	B83501	hypothetical prote
963	6	0.7	242	2	H83829	probable rsbw prot
964	6	0.7	242	2	H70980	ABC-type polar aml
965	6	0.7	243	2	D97008	transcription regu
966	6	0.7	243	1	G86676	hydrogenase (EC 1.
967	6	0.7	244	1	S39400	triosephosphate is
968	6	0.7	244	2	A75408	phosphatase ABC tr
969	6	0.7	244	2	AF2100	1-(5-phosphoribosy
970	6	0.7	245	2	E70544	ketoreductase mmtf
971	6	0.7	246	2	S58176	early auxin-induci
972	6	0.7	246	2	S58497	spou RNA Methylas
973	6	0.7	246	2	G95230	hypothetical prote
974	6	0.7	246	2	G72257	hydroxylase, haloac
975	6	0.7	246	2	A87453	hypothetical prote
976	6	0.7	247	2	G87472	hypothetical prote
977	6	0.7	247	2	T22109	hypothetical prote
978	6	0.7	247	2	D84376	proliferating-cell

687	6	0.7	185	2	C71356	probable ribosomal	760	6	0.7	201	2	G97247	intracellular prot
688	6	0.7	185	2	H95262	hypothetical prote	761	6	0.7	201	2	G48618	hypothetical prote
689	6	0.7	186	2	D86691	methylinphosphotri	762	6	0.7	201	2	G95284	probable WrbA2 trp
690	6	0.7	186	2	JN0517	lumazine protein L	763	6	0.7	202	2	B83059	hypothetical prote
691	6	0.7	186	2	H86623	polypeptide deform	764	6	0.7	202	2	A82160	hypothetical prote
692	6	0.7	186	2	E72000	conserved hypotet	765	6	0.7	202	2	E90663	hypothetical prote
693	6	0.7	186	2	AD0636	RNA polymerase sig	766	6	0.7	202	2	A84325	replication prote
694	6	0.7	186	2	G87589	hypothetical prote	767	6	0.7	202	2	A83492	hypothetical prote
695	6	0.7	186	2	A75281	gag polyprotein -	768	6	0.7	202	2	A85514	probable transcrip
696	6	0.7	187	1	FOVW55	glucanokinase (EC	769	6	0.7	202	2	F82978	guanylate kinase P
697	6	0.7	187	1	S56494	5-formyltetrahydro	770	6	0.7	203	2	S63633	hypothetical prote
698	6	0.7	187	2	H69956	probable lppv prot	771	6	0.7	203	2	T02137	hypothetical prote
699	6	0.7	187	2	C70766	hypothetical prote	772	6	0.7	203	2	F64981	hypothetical 22.4
700	6	0.7	187	2	AE3442	Al4 protein - Indi	773	6	0.7	203	2	F91006	hypothetical prote
701	6	0.7	187	2	S35886	conserved hypotet	774	6	0.7	203	2	G85850	hypothetical prote
702	6	0.7	187	2	A69118	myosin Al catalyti	775	6	0.7	203	2	T08332	conserved hypotet
703	6	0.7	188	2	A23253	myosin light chain	776	6	0.7	204	2	F90483	second ORF in tran
704	6	0.7	189	2	157590	probable scafold1	777	6	0.7	204	2	S76259	hypothetical prote
705	6	0.7	189	2	AE1446	myosin alkali ligh	778	6	0.7	204	2	E70525	hypothetical prote
706	6	0.7	189	2	G96514	peptidylprolyl iso	779	6	0.7	205	2	E82092	4-methyl-5(B)-hydro
707	6	0.7	190	1	MORR1	hypothetical prote	780	6	0.7	205	2	AC0477	hemophore Haas1 m
708	6	0.7	190	2	A11001	hypothetical prote	781	6	0.7	205	2	AC0477	hypothetical prote
709	6	0.7	190	2	T02619	hypothetical prote	782	6	0.7	205	2	E90494	hypothetical prote
710	6	0.7	190	2	D70828	probable 40S ribos	783	6	0.7	205	2	H86675	30S ribosomal prot
711	6	0.7	191	2	S72571	recombinatn prote	784	6	0.7	206	2	A84988	ribosomal prote
712	6	0.7	191	2	A96526	hypothetical prote	785	6	0.7	206	2	B82061	hypothetical prote
713	6	0.7	191	2	T07994	hypothetical prote	786	6	0.7	206	2	T10113	hypothetical prote
714	6	0.7	191	2	C90518	polymyopterin-guan	787	6	0.7	206	2	A05159	peptidylprolyl iso
715	6	0.7	191	2	H87330	hypothetical prote	788	6	0.7	206	2	AE0429	hypothetical prote
716	6	0.7	191	2	E75132	transcription regu	789	6	0.7	206	2	XU1837	methyalted-DNA-pro
717	6	0.7	192	2	A44568	hypothetical prote	790	6	0.7	207	1	XU1837	methyalted-DNA-pro
718	6	0.7	193	2	T44491	hypothetical prote	791	6	0.7	207	1	XU1837	methyalted-DNA-pro
719	6	0.7	193	2	T19081	hypothetical prote	792	6	0.7	207	1	XU1837	methyalted-DNA-pro
720	6	0.7	193	2	F84228	hypothetical prote	793	6	0.7	207	1	XU1837	methyalted-DNA-pro
721	6	0.7	193	2	C75367	hypothetical prote	794	6	0.7	208	1	XU1837	methyalted-DNA-pro
722	6	0.7	193	2	T49977	hypothetical prote	795	6	0.7	209	1	XU1837	methyalted-DNA-pro
723	6	0.7	194	2	C72554	hypothetical prote	796	6	0.7	209	2	AT2662	king-kdpg bifunctio
724	6	0.7	194	2	A64331	hypothetical prote	797	6	0.7	209	2	H97444	dnak -type molecule
725	6	0.7	195	2	S31287	ribosomal protein	798	6	0.7	210	2	C44261	probable hydrolase
726	6	0.7	195	2	A97217	ribosylthrin (lipo	799	6	0.7	210	2	A81955	glyoxalase II fami
727	6	0.7	195	2	S46449	hypothetical prote	800	6	0.7	210	2	B81011	uridine kinase - D
728	6	0.7	195	2	AG0778	probable membrane	801	6	0.7	210	2	E75553	probable acetyltra
729	6	0.7	196	2	I50508	achaeete-scute homo	802	6	0.7	210	2	H75599	dihydrofolate redu
730	6	0.7	196	2	D70467	RNase HII - Aquife	803	6	0.7	211	1	RDBYD	methyalted-DNA-pro
731	6	0.7	196	2	E85553	DNA-3-methyladeni	804	6	0.7	211	1	A41809	dolichyl-phosphate
732	6	0.7	196	2	C72071	hypothetical prote	805	6	0.7	211	2	A75176	sodium-type flagel
733	6	0.7	196	2	G86205	ATP-corrinoid aden	806	6	0.7	211	2	E82055	hypothetical prote
734	6	0.7	196	2	JN0721	COB(1) alamin aden	807	6	0.7	212	2	AH0195	hypothetical prote
735	6	0.7	196	2	AB0654	probable scafold1	808	6	0.7	212	2	F75024	dTMP kinase (EC 2.
736	6	0.7	196	2	D75041	protein W03A5.5 [l	809	6	0.7	212	2	A84359	hypothetical prote
737	6	0.7	196	2	AT1361	aliphatic amidase	810	6	0.7	212	2	T29906	hypothetical prote
738	6	0.7	196	2	G88456	probable transcrip	811	6	0.7	213	2	F86935	probable hypoxanth
739	6	0.7	196	2	B83226	ribosomal protein	812	6	0.7	213	2	B70707	hypothetical prote
740	6	0.7	197	1	T35727	hypothetical prote	813	6	0.7	213	2	AE3487	panthothenate kinas
741	6	0.7	197	1	S16822	hypothetical prote	814	6	0.7	214	2	I49351	granulocyte colony
742	6	0.7	197	2	E71025	hypothetical prote	815	6	0.7	214	2	UC5043	conserved hypotet
743	6	0.7	197	2	F84257	hypothetical prote	816	6	0.7	214	2	H69339	hypothetical prote
744	6	0.7	198	2	T29135	rolp protein - Agr	817	6	0.7	214	2	C75104	hypothetical prote
745	6	0.7	198	2	I39670	thymidine kinase (818	6	0.7	214	2	C71120	transcription regu
746	6	0.7	199	2	D90005	hydrolase family P	819	6	0.7	214	2	AE2934	EpH/GdmH-related
747	6	0.7	199	2	C75363	conserved hypotet	820	6	0.7	215	2	D75342	conserved hpt prote
748	6	0.7	199	2	AT1207	hypothetical prote	821	6	0.7	215	2	A70561	lysine decarboxyla
749	6	0.7	199	2	G83497	conserved hypotet	822	6	0.7	216	2	B97490	conserved hypotet
750	6	0.7	200	2	B95057	hypothetical prote	823	6	0.7	216	2	B97490	conserved hpt prote
751	6	0.7	200	2	E97926	RNA polymerase (de	824	6	0.7	216	2	D89827	conserved hpt prote
752	6	0.7	200	2	S12790	GTP-binding hypotet	825	6	0.7	217	1	F69098	conserved hpt prote
753	6	0.7	200	2	F95985	conserved hypotet	826	6	0.7	217	2	A86846	conserved hpt prote
754	6	0.7	200	2	T34654	probable transmemb	827	6	0.7	217	2	AE3540	phosphate transpor
755	6	0.7	200	2	E83539	hypothetical prote	828	6	0.7	217	2	H64419	H+-transporting tw
756	6	0.7	200	2	T24838	hypothetical prote	829	6	0.7	217	2	B83944	phosphoglycerate m
757	6	0.7	200	2	AC3623	glutaminase (EC 3.	830	6	0.7	217	2	E70574	resolvasase BH2354
758	6	0.7	201	2	AH0664	acyl carrier prote	831	6	0.7	218	2	S30120	probable transcrip
759	6	0.7	201	2	T42647	hypothetical prote	832	6	0.7	218	2	G71875	hypothetical prote

541	6	0.7	152	2	S56198
542	6	0.7	152	2	A86557
543	6	0.7	152	2	F72066
544	6	0.7	152	2	H81675
545	6	0.7	152	2	H83578
546	6	0.7	153	2	J01326
547	6	0.7	153	2	G82555
548	6	0.7	153	2	D81864
549	6	0.7	153	2	D81079
550	6	0.7	153	2	E84218
551	6	0.7	154	2	T36690
552	6	0.7	154	2	H71517
553	6	0.7	154	2	E81984
554	6	0.7	154	2	A81039
555	6	0.7	154	2	AH1951
556	6	0.7	154	2	S66112
557	6	0.7	154	2	AG1021
558	6	0.7	154	2	A99404
559	6	0.7	154	2	G82980
560	6	0.7	154	2	D83516
561	6	0.7	155	2	S38877
562	6	0.7	155	2	E65174
563	6	0.7	155	2	A81015
564	6	0.7	156	2	C72624
565	6	0.7	156	2	I49446
566	6	0.7	156	2	F88175
567	6	0.7	156	2	C84271
568	6	0.7	157	2	B86418
569	6	0.7	158	2	G95186
570	6	0.7	158	2	H98053
571	6	0.7	159	2	H75391
572	6	0.7	159	2	AC2334
573	6	0.7	159	2	B66711
574	6	0.7	159	2	A83219
575	6	0.7	159	2	S72846
576	6	0.7	160	2	G69608
577	6	0.7	160	2	D87481
578	6	0.7	161	2	T39576
579	6	0.7	162	2	A87591
580	6	0.7	162	2	AC0010
581	6	0.7	162	2	B70144
582	6	0.7	163	2	T36985
583	6	0.7	163	2	A75160
584	6	0.7	163	2	D42148
585	6	0.7	164	2	P95026
586	6	0.7	164	2	F97897
587	6	0.7	164	2	C83948
588	6	0.7	164	2	F87438
589	6	0.7	164	2	A97809
590	6	0.7	165	2	H70768
591	6	0.7	165	2	S72810
592	6	0.7	165	2	AH2010
593	6	0.7	165	1	R3B55F
594	6	0.7	166	1	R3B55
595	6	0.7	166	1	S12632
596	6	0.7	166	1	A29240
597	6	0.7	166	1	S12584
598	6	0.7	166	1	T44400
599	6	0.7	166	2	S49101
600	6	0.7	166	2	AB0943
601	6	0.7	166	2	E90471
602	6	0.7	166	2	C33748
603	6	0.7	166	2	S72447
604	6	0.7	167	2	F64953
605	6	0.7	167	2	AH0747
606	6	0.7	167	2	G85803
607	6	0.7	167	2	B90955
608	6	0.7	167	2	AH1975
609	6	0.7	167	2	G86080
610	6	0.7	167	2	G91233
611	6	0.7	167	2	AB1884
612	6	0.7	167	2	AB3204
613	6	0.7	167	2	AB3204

hypothetical prote	614	6	0.7	167	4	S40857	hypothetical prote
CT406 hypothetical	615	6	0.7	168	4	E90403	hypothetical prote
conserved hypotet	616	6	0.7	169	2	I64089	protein-export pro
conserved hypotet	617	6	0.7	169	2	D70226	hypothetical prote
H+-transporting tw	618	6	0.7	170	1	D60010	early E1B 20K prot
c-type cytochrome	619	6	0.7	170	2	AE1391	H+-transporting AT
probable regulator	620	6	0.7	170	2	AG1766	hypothetical prote
regulatory protein	621	6	0.7	170	2	B71132	hypothetical prote
conserved hypotet	622	6	0.7	171	1	XUECAD	hypothetical prote
hypothetical prote	623	6	0.7	171	2	AE0662	hypothetical prote
hypothetical prote	624	6	0.7	171	2	D50868	hypothetical prote
hypothetical prote	625	6	0.7	171	2	E85750	hypothetical prote
conserved hypotet	626	6	0.7	171	2	F87298	hypothetical prote
conserved hypotet	627	6	0.7	171	2	C95143	hypothetical prote
transcription repr	628	6	0.7	171	2	T20567	hypothetical prote
probable acetyltra	629	6	0.7	171	2	A09011	conserved hypotet
hypothetical prote	630	6	0.7	171	2	A46695	epididymal secreto
conserved hypotet	631	6	0.7	172	2	F87000	probable membrane
hypothetical prote	632	6	0.7	172	2	H83139	hypothetical prote
hypothetical prote	633	6	0.7	172	2	H90430	hypothetical prote
troponin C isoform	634	6	0.7	172	2	B86679	hypothetical prote
troponin C isoform	635	6	0.7	172	2	H83542	hypothetical prote
hypothetical 16..7	636	6	0.7	172	2	G0C8AA	hypothetical prote
conserved hypotet	637	6	0.7	173	1	B45932	partial transpos
hypothetical prote	638	6	0.7	173	1	G0GABA	hypothetical prote
8-oxo-dGTPase - mo	639	6	0.7	173	1	D64627	hypothetical prote
protein T24H7.3 [1	640	6	0.7	173	1	JC5982	atrial gland pepti
hypothetical prote	641	6	0.7	173	2	S58632	atrial gland pepti
probable plasma me	642	6	0.7	173	2	AG3476	conserved hypotet
hypothetical prote	643	6	0.7	174	2	E71887	prenylated protein
hypothetical prote	644	6	0.7	174	2	B81148	hypothetical prote
hypothetical prote	645	6	0.7	174	2	T03308	hypothetical prote
hypothetical prote	646	6	0.7	174	2	F96611	hypothetical prote
hypothetical prote	647	6	0.7	174	2	D64150	hypothetical prote
transcription regu	648	6	0.7	174	2	G69824	hypothetical prote
hypothetical prote	649	6	0.7	174	2	A82836	conserved hypotet
transcription fact	650	6	0.7	174	2	D82629	conserved hypotet
hypothetical prote	651	6	0.7	175	2	S03756	hypothetical prote
RNA methylase hom	652	6	0.7	175	2	PW0051	hypothetical prote
conserved hypotet	653	6	0.7	175	2	E69519	synaptojanin 2 del
cytochrome c-type	654	6	0.7	176	2	A84199	conserved hypotet
probable exported	655	6	0.7	176	2	D86434	hypothetical prote
transcription fact	656	6	0.7	176	2	D87345	hypothetical prote
hypothetical prote	657	6	0.7	176	2	C55208	soc3 protein - My
3-isopropylmalate	658	6	0.7	176	2	G71360	hypothetical prote
GTP-binding protei	659	6	0.7	177	2	S45878	hypothetical prote
ribosomal protein	660	6	0.7	177	2	T45667	hypothetical prote
30S ribosomal prot	661	6	0.7	178	2	T13213	hypothetical prote
hypothetical prote	662	6	0.7	178	2	H83061	hypothetical prote
hypothetical prote	663	6	0.7	178	2	P90169	hypothetical prote
hypothetical prote	664	6	0.7	178	2	E88637	hypothetical prote
hypothetical prote	665	6	0.7	178	2	C84723	hypothetical prote
hypothetical prote	666	6	0.7	178	2	D72459	hypothetical prote
hypothetical prote	667	6	0.7	179	2	A87654	hypothetical prote
hypothetical prote	668	6	0.7	179	2	C87652	hypothetical prote
hypothetical prote	669	6	0.7	179	2	C39415	hypothetical prote
hypothetical prote	670	6	0.7	180	2	E48909	hypothetical prote
hypothetical prote	671	6	0.7	180	2	B69524	hypothetical prote
hypothetical prote	672	6	0.7	180	2	F47021	hypothetical prote
hypothetical prote	673	6	0.7	181	2	F84934	hypothetical prote
hypothetical prote	674	6	0.7	181	2	S56532	hypothetical prote
hypothetical prote	675	6	0.7	181	2	T11902	hypothetical prote
hypothetical prote	676	6	0.7	181	2	IVMSB	hypothetical prote
hypothetical prote	677	6	0.7	182	1	AC2175	hypothetical prote
hypothetical prote	678	6	0.7	182	2	E84119	hypothetical prote
hypothetical prote	679	6	0.7	183	2	H87707	hypothetical prote
hypothetical prote	680	6	0.7	183	2	F90354	hypothetical prote
hypothetical prote	681	6	0.7	183	2	I59442	hypothetical prote
hypothetical prote	682	6	0.7	183	2	E90335	hypothetical prote
hypothetical prote	683	6	0.7	184	2	D72307	hypothetical prote
hypothetical prote	684	6	0.7	184	2	S77928	hypothetical prote
hypothetical prote	685	6	0.7	184	2	S59560	hypothetical prote
hypothetical prote	686	6	0.7	185	2		hypothetical prote

hypothetical prote
protein-export pro
hypothetical prote
early E1B 20K prot
H+-transporting AT
H+-transporting AT
hypothetical prote
hypothetical prote
O6-methylguanine-D
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
epididymal secreto
probable membrane
hypothetical prote
partial transpos
hypothetical prote
hypothetical prote
atrial gland pepti
atrial gland pepti
conserved hypotet
prenylated protein
hypothetical prote
methyltransferase
hypothetical prote
hypothetical prote
gene 11 protein, p
hypothetical prote
hypothetical prote
conserved hypotet
conserved hypotet
hypothetical prote
fann protein precu
synaptojanin 2 del
conserved hypotet
hypothetical prote
protein T17H7.16 [1
hypothetical prote
soc3 protein - My
hypothetical prote
hypothetical prote
protein W09G12.6 [1
probable PHD-type
hypothetical prote
hypothetical prote
RNA polymerase s19
fimbrial protein s
G protein-coupled
rRNA (adenine-N6)-
peptic enzyme secr
transcription anti
NADH2 dehydrogenas
interferon beta pr
hypothetical prote
ATP synthase delta
alkyl1 hydrolase
receptor tyrosine
hypothetical prote
conserved hypotet
exoskeletal protei
histone H1.4I - ga

395	6	0.7	107	2	A70966	hypothetical prote
396	6	0.7	107	2	G90307	conserved hypotet
397	6	0.7	108	2	E49898	cellobiose phospho
398	6	0.7	108	2	T13133	protein gp46 - pha
399	6	0.7	108	2	T49144	hypothetical prote
400	6	0.7	108	2	H87647	hypothetical prote
401	6	0.7	109	2	C90768	probable minor tai
402	6	0.7	109	2	E90899	probable minor tai
403	6	0.7	109	2	D90999	probable minor tai
404	6	0.7	109	2	D90997	probable minor tai
405	6	0.7	109	2	C85742	probable tail comp
406	6	0.7	109	2	C85817	hypothetical prote
407	6	0.7	109	2	T42436	FMRFamide-like pep
408	6	0.7	109	2	B72763	hypothetical prote
409	6	0.7	110	2	G65095	hypothetical 12..3
410	6	0.7	110	2	D91123	hypothetical trna synt
411	6	0.7	110	2	C85968	probable trna synt
412	6	0.7	110	2	S74060	hypothetical prote
413	6	0.7	110	2	T00622	hypothetical prote
414	6	0.7	111	2	D72730	hypothetical prote
415	6	0.7	112	4	AS9246	hypothetical prote
416	6	0.7	112	4	SS9290	hypothetical prote
417	6	0.7	113	2	T10469	usp2 protein - Puc
418	6	0.7	113	2	T01657	hypothetical prote
419	6	0.7	113	2	AD1652	hypothetical prote
420	6	0.7	113	2	AG1587	hypothetical prote
421	6	0.7	113	2	T35806	hypothetical prote
422	6	0.7	114	2	H87313	nitrogen regulator
423	6	0.7	114	2	D84852	hypothetical prote
424	6	0.7	114	2	S65224	probable membrane
425	6	0.7	115	2	T04407	probable membrane
426	6	0.7	115	2	S72762	B1496_C2_194 prote
427	6	0.7	115	2	G90123	40S ribosomal prot
428	6	0.7	115	2	H87573	conserved hypotet
429	6	0.7	116	2	T51015	hypothetical prote
430	6	0.7	116	2	F83226	hypothetical prote
431	6	0.7	116	2	T37070	hypothetical prote
432	6	0.7	118	2	I55064	transposase - Esch
433	6	0.7	119	2	D71671	ribosomal protein
434	6	0.7	119	2	A23063	pancreatic B-cell
435	6	0.7	119	2	H83186	hypothetical prote
436	6	0.7	120	2	AE0892	dihydronoepterin a
437	6	0.7	120	2	F83043	hypothetical prote
438	6	0.7	121	2	F84133	hypothetical prote
439	6	0.7	121	2	T22995	hypothetical prote
440	6	0.7	121	2	B87906	protein F59C6.10 l
441	6	0.7	122	1	R5BS12	ribosomal protein
442	6	0.7	123	2	AC3577	probable regulator
443	6	0.7	123	2	D64003	probable sodium-tr
444	6	0.7	124	2	G84953	mutator mult prote
445	6	0.7	124	2	AG5136	conserved hypotet
446	6	0.7	124	2	C69478	NADH dehydrogenase
447	6	0.7	124	2	AC0803	probable bacteriop
448	6	0.7	125	2	E90350	hypothetical prote
449	6	0.7	125	2	AB0928	probable membrane
450	6	0.7	126	2	F87404	ribosomal protein
451	6	0.7	127	2	H69171	hypothetical prote
452	6	0.7	127	2	AS4670	RNA polymerase II
453	6	0.7	127	2	A28084	DNA-binding protei
454	6	0.7	127	2	A70502	probable lprj prot
455	6	0.7	128	2	PC2260	cytochrome P450 pr
456	6	0.7	128	2	AB0070	mutator protein Mu
457	6	0.7	128	2	B69779	conserved hypotet
458	6	0.7	128	2	T48606	hypothetical prote
459	6	0.7	129	2	AD3347	LSU ribosomal prot
460	6	0.7	129	2	S61838	che22 protein - Rh
461	6	0.7	129	2	D86487	14.3k hypothetical
462	6	0.7	130	2	E69384	conserved hypotet
463	6	0.7	131	1	HSZPA3	histone H2A.2 - fl
464	6	0.7	131	1	IRECA2	hypothetical prote
465	6	0.7	131	2	AB0444	4-carboxymuconolac
466	6	0.7	131	2	H75573	hypothetical prote
467	6	0.7	131	2	AB7450	hypothetical prote
<hr/>						
468	6	0.7	132	1	HSZPA2	histone H2A.1 - fl
469	6	0.7	132	2	S24302	hypothetical prote
470	6	0.7	132	2	S00385	hypothetical prote
471	6	0.7	132	2	S69753	hypothetical prote
472	6	0.7	133	2	G90552	competence-damage
473	6	0.7	133	2	AH2405	hypothetical prote
474	6	0.7	133	2	T13212	minor capsid prote
475	6	0.7	134	2	F97475	Id894 (Af322013) l
476	6	0.7	134	2	T45188	hypothetical prote
477	6	0.7	135	1	A61244	hypothetical prote
478	6	0.7	135	2	T36701	natriuretic peptid
479	6	0.7	135	2	F87396	probable transcrip
480	6	0.7	136	2	PQ0517	conserved hypotet
481	6	0.7	137	1	T44983	pol protein - ovin
482	6	0.7	137	2	H75620	methionyl-CoA
483	6	0.7	138	2	B90927	sigma-B regulator
484	6	0.7	138	2	F85775	hypothetical prote
485	6	0.7	138	2	A93662	hypothetical prote
486	6	0.7	139	1	FJECB	transcription term
487	6	0.7	139	2	A24154	85k major surface
488	6	0.7	139	2	E90687	transcription term
489	6	0.7	139	2	A85538	transcription term
490	6	0.7	139	2	S31612	transcription term
491	6	0.7	139	2	T13131	beta-1,3-glucanase
492	6	0.7	139	2	H87428	protein gp45 - pha
493	6	0.7	140	2	H71033	conserved hypotet
494	6	0.7	140	2	S73701	hypothetical prote
495	6	0.7	140	2	T25664	hypothetical prote
496	6	0.7	141	2	G75455	thioredoxin - Dein
497	6	0.7	141	2	S55247	hemoglobin alpha-A
498	6	0.7	142	2	AC0928	probable regulator
499	6	0.7	143	1	E64901	protein C, osmotic
500	6	0.7	143	2	D72785	yeast homolog Ape02
501	6	0.7	143	2	AD0673	osmotically induci
502	6	0.7	143	2	C85728	osmotically induci
503	6	0.7	143	2	F90889	osmotically induci
504	6	0.7	143	2	T28838	hypothetical prote
505	6	0.7	145	2	S54230	hypothetical prote
506	6	0.7	145	2	D83444	Ig mu heavy chain
507	6	0.7	145	2	D64323	conserved hypotet
508	6	0.7	145	2	D69043	hypothetical prote
509	6	0.7	145	2	T00967	hypothetical prote
510	6	0.7	146	2	E81977	hypothetical prote
511	6	0.7	146	2	F81032	probable acetyltra
512	6	0.7	146	2	A86490	ribosomal-protein-
513	6	0.7	146	2	H83244	protein T32E20.11
514	6	0.7	146	2	T36621	hypothetical prote
515	6	0.7	146	2	A47264	hypothetical prote
516	6	0.7	147	1	S36122	sRNAP core protein
517	6	0.7	147	2	G72691	phosphotransferase
518	6	0.7	147	2	D85950	probable glycine C
519	6	0.7	147	2	A98105	hypothetical prote
520	6	0.7	147	2	A70528	hypothetical prote
521	6	0.7	147	2	B97581	hypothetical prote
522	6	0.7	147	2	AH2801	mult like protein
523	6	0.7	148	2	T39518	40S ribosomal prot
524	6	0.7	148	2	H70742	hypothetical prote
525	6	0.7	148	2	E90306	conserved hypotet
526	6	0.7	148	2	D95130	hypothetical prote
527	6	0.7	148	2	C98001	hypothetical prote
528	6	0.7	148	2	B84510	hypothetical prote
529	6	0.7	149	2	S23668	hypothetical prote
530	6	0.7	149	2	B69488	ccl2 protein - Rho
531	6	0.7	149	2	T47597	LSU ribosomal prot
532	6	0.7	149	2	AH2262	hypothetical prote
533	6	0.7	150	1	MORFA2	hypothetical prote
534	6	0.7	150	2	S54791	myosin alkali ligh
535	6	0.7	150	2	B23253	superoxide dismuta
536	6	0.7	150	2	H64576	myosin A2 catalyti
537	6	0.7	150	2	D87463	hypothetical prote
538	6	0.7	150	2	E81332	hypothetical prote
539	6	0.7	151	2	E81310	sepd protein - Esc
540	6	0.7	152	2	B82389	transcription regu

249	7	0.8	875	2	D81651	conserved hypoteth	322	6	0.7	72	2	E97706	hypothetical prote
250	7	0.8	947	2	E86362	hypothetical prote	323	6	0.7	73	2	T31216	hypothetical prote
251	7	0.8	949	2	D90803	Aida-1 adhesin-lik	324	6	0.7	74	2	S28794	H+-transporting tw
252	7	0.8	987	2	T40241	probable guanin n	325	6	0.7	74	2	S63632	H+-transporting tw
253	7	0.8	992	2	T38817	hypothetical prote	326	6	0.7	74	2	E87294	AMP synthase F0, C
254	7	0.8	993	2	A10669	probable virulence	327	6	0.7	74	2	G85899	probable oxidoredu
255	7	0.8	996	2	E98200	sarcosine oxidase	328	6	0.7	76	2	S26943	H+-transporting tw
256	7	0.8	996	2	AD3086	sarcosine oxidase	329	6	0.7	76	2	B96809	protein F28K19.18
257	7	0.8	1005	2	H85611	probable adhesin Z	330	6	0.7	77	2	JH0649	conserved hypoteth
258	7	0.8	1026	2	G81751	exodeoxyribonuclea	331	6	0.7	78	1	G69508	reverse transcript
259	7	0.8	1026	1	A43322	1-phosphatidylinos	332	6	0.7	78	2	T03670	p83/100 protein -
260	7	0.8	1068	1	T38110	1-phosphatidylinos	333	6	0.7	78	2	S61473	p83/100 protein -
261	7	0.8	1134	2	S53955	hypothetical prote	334	6	0.7	78	2	S61475	p83/100 protein -
262	7	0.8	1157	2	E97255	fusion of alpha-g1	335	6	0.7	78	2	S61474	hypothetical prote
263	7	0.8	1223	2	S62011	PH085 protein - ye	336	6	0.7	79	2	C87028	hypothetical prote
264	7	0.8	1228	2	T40468	surface layer prot	337	6	0.7	79	2	H70827	hypothetical prote
265	7	0.8	1252	2	S36016	ococyst wall protei	338	6	0.7	79	2	AC0013	reverse transcript
266	7	0.8	1259	2	A43425	Bravo/Nr-CAM cell	339	6	0.7	81	2	T03723	reverse transcript
267	7	0.8	1268	1	A39640	neural cell adhesi	340	6	0.7	81	2	T03707	reverse transcript
268	7	0.8	1286	1	T02187	probable ABC trans	341	6	0.7	81	2	AC0142	conserved hypoteth
269	7	0.8	1292	2	T48007	P-glycoprotein hom	342	6	0.7	82	2	H83584	probable biotin-re
270	7	0.8	1300	2	T53799	CG1 protein - huma	343	6	0.7	82	2	AF3441	hypothetical prote
271	7	0.8	1302	2	C81182	iron-regulated pro	344	6	0.7	83	2	AF3441	hypothetical prote
272	7	0.8	1322	2	T01842	xanthine dehydroge	345	6	0.7	86	2	C81052	acyl carrier prote
273	7	0.8	1324	2	S07245	xanthine dehydroge	346	6	0.7	86	2	S6073	veg protein - Bac
274	7	0.8	1342	2	A31946	kinectin 1 - human	347	6	0.7	86	2	C97450	exodeoxyribonuclea
275	7	0.8	1356	2	S32763	collagen alpha 1(X	348	6	0.7	86	2	AE2668	conserved hypoteth
276	7	0.8	1388	2	A53317	hypothetical prote	349	6	0.7	87	2	AF0910	ccg-4 protein 2 (1
277	7	0.8	1434	2	G71232	protein F17L21.22	350	6	0.7	87	2	T47209	1beta-hydroxyster
278	7	0.8	1475	2	F86399	hypothetical prote	351	6	0.7	88	2	S60188	hypothetical prote
279	7	0.8	1539	2	T30037	cytotoxin RTX homo	352	6	0.7	88	2	C83657	hypothetical prote
280	7	0.8	1829	2	S35027	iron-regulated pro	353	6	0.7	88	2	AG1356	hypothetical prote
281	7	0.8	1829	2	E81086	MHC class III hist	354	6	0.7	88	2	AH1726	hypothetical prote
282	7	0.8	1870	2	S37671	MHC class III hist	355	6	0.7	89	2	A84301	hypothetical prote
283	7	0.8	1872	2	S36152	MHC class III hist	356	6	0.7	90	2	T28989	2'-aminobiphenyl-2
284	7	0.8	1886	2	S04921	nuclear pore prote	357	6	0.7	90	2	T46580	probable cysteine
285	7	0.8	1914	2	T42635	tenascin Y precurs	358	6	0.7	91	2	S52275	hypothetical prote
286	7	0.8	1939	2	AF0095	probable sideropho	359	6	0.7	91	2	D75561	hypothetical prote
287	7	0.8	2142	2	B35098	MHC class III hist	360	6	0.7	91	2	H97856	hypothetical prote
288	7	0.8	2336	2	A45386	omega-conotoxin-se	361	6	0.7	92	2	H84101	hypothetical prote
289	7	0.8	2698	2	B96671	similar to transa	362	6	0.7	92	2	B90684	unknown [imported]
290	7	0.8	3157	2	B70969	probable PPE prote	363	6	0.7	93	2	S37677	probable membrane
291	7	0.8	3415	2	A46105	polyprotein(NS1, N	364	6	0.7	93	2	F85534	hypothetical prote
292	7	0.8	3716	2	E70969	probable PPE prote	365	6	0.7	93	2	G64886	hypothetical prote
293	7	0.8	7829	2	T15789	hypothetical prote	366	6	0.7	93	2	G64886	hypothetical prote
294	6	0.7	12	2	S29830	dimethylalliline mo	367	6	0.7	93	2	H64767	hypothetical prote
295	6	0.7	22	2	S05236	exoenzyme C3 - Clo	368	6	0.7	94	2	B72867	hypothetical prote
296	6	0.7	25	2	T09385	rev protein - huma	369	6	0.7	94	2	S75465	AuAI protein - yea
297	6	0.7	32	2	T17394	vin protein - DIC	370	6	0.7	94	2	S32939	calpactin I light
298	6	0.7	33	2	A56818	Na+/K+-exchanging	371	6	0.7	95	1	L0PG10	ferredoxin (2Fe-2S
299	6	0.7	42	2	S7815	probable DNA ligas	372	6	0.7	96	1	FEAH	hypothetical prote
300	6	0.7	51	2	S01870	hypothetical prote	373	6	0.7	96	2	AB3053	aggregran chondroli
301	6	0.7	52	2	S44790	D2007.3 protein -	374	6	0.7	97	2	I51262	calpactin I light
302	6	0.7	52	2	G82835	hypothetical prote	375	6	0.7	97	2	JC1139	calpactin I light
303	6	0.7	52	2	B98004	hypothetical prote	376	6	0.7	97	2	B28489	hypothetical prote
304	6	0.7	55	2	AH2396	hemolymph trypsin	377	6	0.7	98	2	C86517	hypothetical prote
305	6	0.7	56	2	A29235	30S ribosomal prot	378	6	0.7	98	2	G84296	hypothetical prote
306	6	0.7	57	2	C84213	hypothetical prote	379	6	0.7	98	2	D72106	Alzheimer's diseas
307	6	0.7	60	2	T03816	hypothetical prote	380	6	0.7	100	2	A32282	gene Pax-1 protein
308	6	0.7	61	2	S60726	M protein precurs	381	6	0.7	100	2	F72530	conserved hypoteth
309	6	0.7	61	2	D69526	conserved hypoteth	382	6	0.7	102	2	I53094	conserved hypoteth
310	6	0.7	62	2	PK0018	GTP-binding protel	383	6	0.7	102	2	A12693	conserved hypoteth
311	6	0.7	63	2	C96942	hypothetical prote	384	6	0.7	104	2	G89574	Met regulon regula
312	6	0.7	64	2	B82672	hypothetical prote	385	6	0.7	105	1	RGECEW	repressor of all m
313	6	0.7	66	2	S77700	thiol protein - Es	386	6	0.7	105	2	C91237	transferrin complex p
314	6	0.7	66	2	H86090	thiamin biosynthes	387	6	0.7	105	2	G86084	hypothetical prote
315	6	0.7	68	2	AF0066	hypothetical prote	388	6	0.7	105	2	T44766	hypothetical prote
316	6	0.7	69	2	S70158	hypothetical prote	389	6	0.7	105	2	G89993	hypothetical prote
317	6	0.7	69	2	F90926	hypothetical prote	390	6	0.7	106	2	T36932	probable integral
318	6	0.7	69	2	B85775	hypothetical prote	391	6	0.7	106	2	AG0216	conserved hypoteth
319	6	0.7	69	2	C64925	hypothetical prote	392	6	0.7	106	2	F87608	hypothetical prote
320	6	0.7	70	2	AH0700	conserved hypoteth	393	6	0.7	106	2	AI2219	hypothetical prote
321	6	0.7	71	2	T50483	glucose-repressibl	394	6	0.7	106	2		

103	7	0.8	323	2	AH0903	octaprenyl-diphosp
104	7	0.8	323	2	AC2119	hypothetical prote
105	7	0.8	331	2	T02343	glucan endo-1,3-be
106	7	0.8	337	2	A12360	hypothetical prote
107	7	0.8	338	2	SA5078	probable membrane
108	7	0.8	340	2	A12922	hypothetical prote
109	7	0.8	341	2	C89789	hypothetical prote
110	7	0.8	343	2	D83282	probable permease
111	7	0.8	344	2	B97697	sugar ABC transpor
112	7	0.8	347	2	A65022	hypothetical prote
113	7	0.8	347	2	H91044	hypothetical prote
114	7	0.8	347	2	C85889	hypothetical prote
115	7	0.8	351	2	T44428	probable gonococci
116	7	0.8	352	2	E82263	conserved hypotbet
117	7	0.8	355	2	D84192	glutamate decarbox
118	7	0.8	356	2	T00881	probable PCF2-like
119	7	0.8	358	2	H87267	heat-inducible tra
120	7	0.8	368	2	T36414	probable iron-side
121	7	0.8	368	2	AC1508	homoserine O-acety
122	7	0.8	368	2	AC1508	homoserine O-acety
123	7	0.8	371	2	B83781	two-component sens
124	7	0.8	375	1	A23689	limulus clotting e
125	7	0.8	379	2	F72022	hypothetical prote
126	7	0.8	379	2	H86600	hypothetical prote
127	7	0.8	382	2	B82177	Na+/H+ antiporter
128	7	0.8	382	2	D91038	hypothetical prote
129	7	0.8	382	2	G91276	hypothetical prote
130	7	0.8	383	2	JY0360	Na+/H+-exchanging
131	7	0.8	383	2	A70612	hypothetical prote
132	7	0.8	385	2	A83466	probable RND efflu
133	7	0.8	389	2	T22465	hypothetical prote
134	7	0.8	389	2	H82825	transport protein
135	7	0.8	395	2	S74950	patal protein - Sy
136	7	0.8	396	2	G69808	multidrug resistan
137	7	0.8	397	2	T35609	while protein I - S
138	7	0.8	398	2	A37274	glia-derived nexin
139	7	0.8	401	2	AC2113	alanine racemase I
140	7	0.8	402	2	C64895	hypothetical prote
141	7	0.8	402	2	G85882	probable virulence
142	7	0.8	402	2	G86117	probable IS elemen
143	7	0.8	402	2	AH0514	hypothetical prote
144	7	0.8	402	2	T25001	coenzyme F420 hydr
145	7	0.8	405	2	A35620	triglycer factor tig
146	7	0.8	424	2	F69723	deoxyguanosinetrip
147	7	0.8	426	2	D97566	cytochrome P450 Rv
148	7	0.8	426	2	AB2787	hypothetical prote
149	7	0.8	428	1	F70729	probable permease
150	7	0.8	430	2	T46099	probable uracil tr
151	7	0.8	432	2	AF0697	probable transport
152	7	0.8	435	2	G90700	probable membrane
153	7	0.8	435	2	B85551	probable deaminase
154	7	0.8	441	2	H64782	hypothetical prote
155	7	0.8	441	2	D95390	hypothetical prote
156	7	0.8	451	2	B96495	probable phosphom
157	7	0.8	454	2	E70637	probable phosphom
158	7	0.8	465	2	C70594	M20/M25/M40 family
159	7	0.8	471	2	E87340	RuvB DNA helicase-
160	7	0.8	473	2	T46049	exodeoxyribonuclea
161	7	0.8	477	2	C81865	hypothetical prote
162	7	0.8	480	2	E83105	alpha,alpha-trehal
163	7	0.8	482	2	E70554	hypothetical prote
164	7	0.8	488	2	S38987	hypothetical prote
165	7	0.8	496	2	T50272	probable aldehyde
166	7	0.8	496	2	A56040	hypothetical prote
167	7	0.8	496	2	A56040	hypothetical prote
168	7	0.8	499	2	A12449	protein-tyrosine k
169	7	0.8	505	2	C96300	hypothetical prote
170	7	0.8	505	2	AD2983	hypothetical prote
171	7	0.8	507	1	C53308	probable hydrolase
172	7	0.8	512	1	ALBSL	alpha-amylase (EC
173	7	0.8	512	2	G91200	type III secretion
174	7	0.8	512	2	T80311	sepc protein - Esc
175	7	0.8	512	2	C86047	escC [imported] -
176	7	0.8	513	2	S54469	hypothetical prote
177	7	0.8	514	2	A87266	sensor histidine k
178	7	0.8	516	2	F86327	protein F18014.21
179	7	0.8	518	2	AB3566	type I restriction
180	7	0.8	519	1	A54549	glucan 1,4-alpha-g
181	7	0.8	519	2	B45697	immediate-early pr
182	7	0.8	521	2	S02018	regulatory protein
183	7	0.8	522	2	E82024	peptide methionine
184	7	0.8	522	2	G81243	peptide methionine
185	7	0.8	523	2	H70886	probable spermidin
186	7	0.8	528	2	A70915	probable PE protei
187	7	0.8	533	2	D71202	hypothetical prote
188	7	0.8	533	2	A70915	probable ATP-bind
189	7	0.8	537	1	A64085	hypothetical prote
190	7	0.8	559	2	B72487	hypothetical prote
191	7	0.8	566	2	D84611	probable cytochrom
192	7	0.8	566	2	A72254	probable cytochrom
193	7	0.8	567	2	T33535	methy1-accepting c
194	7	0.8	567	2	D83400	Ras pathway intera
195	7	0.8	567	2	AB0006	conserved hypotbet
196	7	0.8	571	2	T39378	probable DNA ligas
197	7	0.8	574	2	E65135	hypothetical 64.6K
198	7	0.8	574	2	B87619	sensor histidine k
199	7	0.8	579	2	A96499	cerebroglycan prec
200	7	0.8	584	2	B90281	hypothetical prote
201	7	0.8	584	2	T14631	glucose-6-phosphat
202	7	0.8	587	2	F83336	inner membrane cop
203	7	0.8	595	1	A48077	myb-related protei
204	7	0.8	597	2	T98186	adenine deaminase
205	7	0.8	597	2	AF3100	adenine deaminase
206	7	0.8	612	2	B82756	beta-galactosidase
207	7	0.8	613	2	F82374	dihydroxy-acid deh
208	7	0.8	614	2	A95072	exonuclease ABC,
209	7	0.8	617	2	G97939	conserved hypotbet
210	7	0.8	617	2	G82706	outer membrane pro
211	7	0.8	623	2	A13488	hypothetical prote
212	7	0.8	627	2	T04562	numb-binding prote
213	7	0.8	628	2	T09458	Rppp-related prot
214	7	0.8	629	2	T48799	hypothetical prote
215	7	0.8	629	2	T34370	unknown protein F1
216	7	0.8	630	2	G96656	acyl-CoA oxidase h
217	7	0.8	639	2	T13151	hypothetical prote
218	7	0.8	641	2	T44585	conserved hypotbet
219	7	0.8	652	2	B83725	transporter BME107
220	7	0.8	679	2	D69329	DNA ligase - Therm
221	7	0.8	682	2	AG3351	synthaxin-binding p
222	7	0.8	688	2	D72418	probable transcrip
223	7	0.8	693	2	T41443	hypothetical prote
224	7	0.8	695	2	B95879	hypothetical prote
225	7	0.8	698	2	A69222	numb-binding prote
226	7	0.8	706	2	G72499	conserved secreted
227	7	0.8	728	2	T09457	hypothetical prote
228	7	0.8	732	2	AB2732	hypothetical prote
229	7	0.8	741	2	B97513	probable secreted
230	7	0.8	744	1	IS1022	hypothetical prote
231	7	0.8	745	2	T03119	hypothetical prote
232	7	0.8	752	2	G82798	hypothetical prote
233	7	0.8	767	2	T05662	hypothetical prote
234	7	0.8	776	2	T15411	hypothetical prote
235	7	0.8	795	1	S22804	glycogen debranchl
236	7	0.8	804	2	H75549	replication licens
237	7	0.8	807	1	IS1685	replication licens
238	7	0.8	808	1	S62594	replication licens
239	7	0.8	815	2	T00538	probable serine pro
240	7	0.8	817	2	T16409	hypothetical prote
241	7	0.8	830	2	F83288	hypothetical prote
242	7	0.8	830	2	S25198	conserved hypotbet
243	7	0.8	839	2	T45908	vacuolar membrane
244	7	0.8	846	2	AD0279	hypothetical prote
245	7	0.8	850	2	AF3484	probable virulence
246	7	0.8	862	2	G87094	hypothetical trans
247	7	0.8	862	2	E84567	probable aminopept
248	7	0.8	862	2	T01141	probable trehalose
						hypothetical prote

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:32:03 ; Search time 31 Seconds
(without alignments)
2849,920 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 919
Sequence: 1 MSKPVLPANRSEMPVALAA.....TGKKEGCKPIKLFHFICTPF 919

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_73:*

1: pir1:*\n2: pir2:*\n3: pir3:*\n4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	1.4	578	2	AE0428
2	11	1.2	577	2	AD0455
3	11	1.2	578	2	B64012
4	10	1.1	579	2	F83327
5	9	1.0	851	2	A46160
6	8	0.9	174	2	A70651
7	8	0.9	200	2	H64989
8	8	0.9	200	2	C91015
9	8	0.9	200	2	E85859
10	8	0.9	331	2	AE1390
11	8	0.9	331	2	AG1765
12	8	0.9	340	2	T44739
13	8	0.9	360	2	T27569
14	8	0.9	380	2	JC5747
15	8	0.9	469	2	T34645
16	8	0.9	577	2	S36445
17	8	0.9	577	2	F91278
18	8	0.9	577	2	F86119
19	8	0.9	582	2	F82064
20	8	0.9	731	2	AH2546
21	8	0.9	1355	2	S40022
22	8	0.9	1402	2	S42748
23	8	0.9	1417	2	T00661
24	7	0.8	66	2	G82667
25	7	0.8	72	2	S75519
26	7	0.8	82	2	G55116
27	7	0.8	82	2	AG0884
28	7	0.8	82	2	C85958
29	7	0.8	82	2	C91113

30	7	0.8	88	2	AH0088	probable flagellar
31	7	0.8	91	2	A81784	hypothetical prote
32	7	0.8	95	2	T18004	hypothetical prote
33	7	0.8	107	2	S61135	hypothetical prote
34	7	0.8	109	2	S73639	M6353 homolog G12-
35	7	0.8	119	1	A69441	conserved hypotnet
36	7	0.8	119	2	A97825	50S ribosomal prot
37	7	0.8	124	2	A48075	60K stress-related
38	7	0.8	129	2	F97593	50S ribosomal prot
39	7	0.8	129	2	AC2815	hypothetical prote
40	7	0.8	130	2	AC1545	hypothetical prote
41	7	0.8	130	2	AD1187	hypothetical prote
42	7	0.8	136	2	CG9892	Dnak suppressor pr
43	7	0.8	138	2	F81243	probable Dnak supp
44	7	0.8	138	2	G82015	hypothetical prote
45	7	0.8	144	2	F72556	hypothetical prote
46	7	0.8	157	2	A84402	hypothetical prote
47	7	0.8	159	2	B70635	hypothetical prote
48	7	0.8	164	2	G72262	conserved hypotnet
49	7	0.8	166	2	D71049	probable I(+) -tart
50	7	0.8	174	2	CG4174	hypothetical prote
51	7	0.8	177	2	F96739	hypothetical prote
52	7	0.8	177	2	E83545	hypothetical prote
53	7	0.8	179	2	S59503	ferric pseudobactl
54	7	0.8	190	2	A33839	23K calcium-bindin
55	7	0.8	194	2	C96740	hypothetical prote
56	7	0.8	195	1	IVBO11	interferon alpha-I
57	7	0.8	202	2	F82590	hypothetical prote
58	7	0.8	206	2	G97061	hypothetical prote
59	7	0.8	211	1	A40836	formate dehydrogen
60	7	0.8	211	2	A86078	formate dehydrogen
61	7	0.8	211	2	B91231	formate dehydrogen
62	7	0.8	211	2	AD0946	formate dehydrogen
63	7	0.8	211	2	S15167	fix protein - Azo
64	7	0.8	212	2	G75458	conserved hypotnet
65	7	0.8	213	2	A95013	hypothetical prote
66	7	0.8	213	2	D97884	hypothetical prote
67	7	0.8	214	2	T21585	hypothetical prote
68	7	0.8	214	2	G81237	hypothetical prote
69	7	0.8	225	2	C95082	amino acid ABC tra
70	7	0.8	225	2	G97949	phosphoglycolate p
71	7	0.8	226	2	F82052	hypothetical prote
72	7	0.8	231	2	C83015	hypothetical prote
73	7	0.8	232	2	S26023	hypothetical prote
74	7	0.8	239	2	H82010	hypothetical prote
75	7	0.8	242	2	B64917	transcription regu
76	7	0.8	242	2	B90918	hypothetical prote
77	7	0.8	242	2	G85766	hypothetical prote
78	7	0.8	243	2	AC0690	probable two-compo
79	7	0.8	245	2	AB0499	ABC transporter pe
80	7	0.8	251	2	C72768	probable electon
81	7	0.8	256	2	AG2545	hypothetical prote
82	7	0.8	266	2	F70595	hypothetical prote
83	7	0.8	271	2	C86907	regulator of purin
84	7	0.8	277	1	F69398	dihydroxynaphthoic
85	7	0.8	280	2	A49921	conserved hypotnet
86	7	0.8	281	2	A43749	lignin beta-ether
87	7	0.8	282	2	E93354	hypothetical prote
88	7	0.8	285	2	H69369	branched-chain ami
89	7	0.8	293	2	C71267	probable carboxyle
90	7	0.8	294	2	C64048	major ferric iron-
91	7	0.8	297	2	JC4398	thiosulfate sulfur
92	7	0.8	297	2	JC5286	thiosulfate sulfur
93	7	0.8	298	2	CG1516	hypothetical prote
94	7	0.8	303	2	H70716	hypothetical prote
95	7	0.8	305	2	C97405	conserved hypotnet
96	7	0.8	305	2	AC2623	hypothetical prote
97	7	0.8	310	2	F86737	hypothetical prote
98	7	0.8	313	2	AG1656	glycerate dehydrog
99	7	0.8	317	2	S51572	moca protein - Rhl
100	7	0.8	323	2	E65109	octaprenyl -diphosp
101	7	0.8	323	2	B91137	octaprenyl diphosp
102	7	0.8	323	2	E85982	octaprenyl diphosp

```

; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C31
; CURRENT APPLICATION NUMBER: US/10/174,579
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-579-356

```

```

Query Match          0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 404 LOAVRAL 410
Db 76 LOAVRAL 82

```

```

RESULT 40
US-10-174-582-356
; Sequence 356, Application US/10174582
; Publication No. US20030027265A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C36
; CURRENT APPLICATION NUMBER: US/10/174,582
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-582-356

```

```

Query Match          0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 404 LOAVRAL 410
Db 76 LOAVRAL 82

```

```

Search completed: April 28, 2003, 16:41:50
Job time : 88 secs

```

```
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C13
; CURRENT APPLICATION NUMBER: US/10/180,552
; CURRENT FILING DATE: 2002-06-25
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-180-552-356
```

```
Query Match 0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 404 LQAVRAL 410
    |||||||
DB 76 LQAVRAL 82
```

```
RESULT 36
US-10-180-557-356
; Sequence 356, Application US/10180557
; Publication No. US20030022301A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C147
; CURRENT APPLICATION NUMBER: US/10/180,557
; CURRENT FILING DATE: 2002-06-25
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-180-557-356
```

```
Query Match 0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 404 LQAVRAL 410
    |||||||
DB 76 LQAVRAL 82
```

```
RESULT 37
US-10-173-700-356
; Sequence 356, Application US/10173700
; Publication No. US20030027262A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
```

```
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C14
; CURRENT APPLICATION NUMBER: US/10/173,700
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-173-700-356
```

```
Query Match 0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 404 LQAVRAL 410
    |||||||
DB 76 LQAVRAL 82
```

```
RESULT 38
US-10-174-572-356
; Sequence 356, Application US/10174572
; Publication No. US20030027263A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C40
; CURRENT APPLICATION NUMBER: US/10/174,572
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-174-572-356
```

```
Query Match 0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 404 LQAVRAL 410
    |||||||
DB 76 LQAVRAL 82
```

```
RESULT 39
US-10-174-579-356
; Sequence 356, Application US/10174579
; Publication No. US20030027264A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
```

```

; FILE REFERENCE: P3430RIC60
; CURRENT APPLICATION NUMBER: US/10/175,752
; CURRENT FILING DATE: 2002-06-19
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430RIC60
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-752-356

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 404 LQAVRAL 410
DB 76 LQAVRAL 82

```

```

RESULT 32
US-10-176-482-356
; Sequence 356, Application US/10176482
; Publication No. US20030022296A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC70
; CURRENT APPLICATION NUMBER: US/10/176,482
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-482-356

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 404 LQAVRAL 410
DB 76 LQAVRAL 82

```

```

RESULT 33
US-10-176-757-356
; Sequence 356, Application US/10176757
; Publication No. US20030022297A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.

```

```

; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC60
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-757-356

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 404 LQAVRAL 410
DB 76 LQAVRAL 82

```

```

RESULT 34
US-10-176-913-356
; Sequence 356, Application US/10176913
; Publication No. US20030022298A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC66
; CURRENT APPLICATION NUMBER: US/10/176,913
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-913-356

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 404 LQAVRAL 410
DB 76 LQAVRAL 82

```

```

RESULT 35
US-10-180-552-356
; Sequence 356, Application US/10180552
; Publication No. US20030022300A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James

```

```
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-356
```

```
Query Match          0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 404 LQAVRAL 410
      |||||
Db 76 LQAVRAL 82
```

RESULT 28

```
US-10-063-502-118
; Sequence 118, Application US/10063502
; Publication No. US20030023042A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,502
; FILE REFERENCE: P3430R1C45
; CURRENT FILING DATE: 2002-05-01
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 118
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-502-118
```

```
Query Match          0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 404 LQAVRAL 410
      |||||
Db 76 LQAVRAL 82
```

RESULT 29

```
US-10-173-706-356
; Sequence 356, Application US/10173706
; Publication No. US20030022293A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C7
; CURRENT APPLICATION NUMBER: US/10/173,706
; CURRENT FILING DATE: 2002-06-17
```

```
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-706-356
```

```
Query Match          0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 404 LQAVRAL 410
      |||||
Db 76 LQAVRAL 82
```

RESULT 30

```
US-10-175-738-356
; Sequence 356, Application US/10175738
; Publication No. US20030022294A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C45
; CURRENT APPLICATION NUMBER: US/10/175,738
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-738-356
```

```
Query Match          0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 404 LQAVRAL 410
      |||||
Db 76 LQAVRAL 82
```

RESULT 31

```
US-10-175-752-356
; Sequence 356, Application US/10175752
; Publication No. US20030022295A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
```

; ORGANISM: Homo Sapien
US-10-063-547-118

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410
|||||||
DB 76 LQAVRAL 82

RESULT 24

US-10-174-590-356
; Sequence 356, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-356

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410
|||||||
DB 76 LQAVRAL 82

RESULT 25

US-10-176-758-356
; Sequence 356, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; CURRENT FILING DATE: 2002-06-21
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-356

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410
|||||||
DB 76 LQAVRAL 82

RESULT 26

US-10-063-616-118
; Sequence 118, Application US/10063616
; Publication No. US20030013855A1
; GENERAL INFORMATION:
; APPLICANT: Baton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,616
; CURRENT FILING DATE: 2002-05-03
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 118
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-616-118

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410
|||||||
DB 76 LQAVRAL 82

RESULT 27

US-10-175-737-356
; Sequence 356, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175,737
; CURRENT FILING DATE: 2002-06-19
; Prior application removed - See File Wrapper or Palm

SEQ ID NO 162
LENGTH: 188
TYPE: PRT
ORGANISM: Glycine max
US-10-062-254-162

Query Match
Best Local Similarity 100.0%; Score 7; DB 12; Length 188;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 LITVNMG 399
|||||||
DB 46 LITVNMG 52

RESULT 20
US-09-764-868-956
Sequence 956, Application US/09764868
Patent No. US2002016871A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT232
CURRENT APPLICATION NUMBER: US/09/764,868
PRIORITY FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 1510
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 956
LENGTH: 214
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (183)
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (204)
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (209)
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-764-868-956

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 214;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 DRYLAIN 514
|||||||
DB 106 DRYLAIN 112

RESULT 21
US-09-825-414-60
Sequence 60, Application US/09825414
Patent No. US20020083489A1
GENERAL INFORMATION:
APPLICANT: Collier, Alan
APPLICANT: Altano, James R.
APPLICANT: Charkowski, Amy O.
TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES
FILE REFERENCE: 19603/3743
CURRENT APPLICATION NUMBER: US/09/825,414
PRIORITY FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/194,160
PRIOR FILING DATE: 2000-04-03
PRIOR APPLICATION NUMBER: 60/224,604
PRIOR FILING DATE: 2000-08-11
PRIOR APPLICATION NUMBER: 60/249,548
PRIOR FILING DATE: 2000-11-17
NUMBER OF SEQ ID NOS: 91

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 60
LENGTH: 215
TYPE: PRT
ORGANISM: Pseudomonas syringae pv. delphinii
US-09-825-414-60

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 215;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 IRGYAHD 825
|||||||
DB 146 IRGYAHD 152

RESULT 22
US-09-866-050A-705
Sequence 705, Application US/09866050A
Publication No. US20030040471A1
GENERAL INFORMATION:
APPLICANT: Watson, James D.
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Onrust, Rene
APPLICANT: Murison, James G.
APPLICANT: Kumble, Krishanand D.
TITLE OF INVENTION: Compositions and Methods for Their Use
TITLE OF INVENTION: and Methods for Their Use
FILE REFERENCE: 11000,1011c4U
CURRENT APPLICATION NUMBER: US/09/866,050A
PRIORITY FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 725
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 705
LENGTH: 255
TYPE: PRT
ORGANISM: Mouse
US-09-866-050A-705

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 255;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410
|||||||
DB 76 LQAVRAL 82

RESULT 23
US-10-063-547-118
Sequence 118, Application US/10063547
Publication No. US20020182638A1
GENERAL INFORMATION:
APPLICANT: Eaton, Dan L.
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C1
CURRENT APPLICATION NUMBER: US/10/063,547
PRIORITY FILING DATE: 2002-05-02
Prior application removed - see file wrapper or Palm
NUMBER OF SEQ ID NOS: 170
SEQ ID NO 118
LENGTH: 261
TYPE: PRT

QY 166 KRLYARL 172
Db 5 KRLYARL 11

RESULT 16
US-09-975-132A-27
; Sequence 27, Application US/09975132A
; Publication No. US20020182672A1
; GENERAL INFORMATION:
; APPLICANT: Kolman, Marc
; TITLE OF INVENTION: Enhanced Secretion of a Polypeptide by a
; FILE REFERENCE: GC636-2
; CURRENT APPLICATION NUMBER: US/09/975,132A
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US 60/239,531
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: IL-3 substituted tag changed C-terminus
US-09-975-132A-27

Query Match 0.8%; Score 7; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 KRLYARL 172
Db 5 KRLYARL 11

RESULT 17
US-09-975-132A-28
; Sequence 28, Application US/09975132A
; Publication No. US20020182672A1
; GENERAL INFORMATION:
; APPLICANT: Kolman, Marc
; TITLE OF INVENTION: Enhanced Secretion of a Polypeptide by a
; FILE REFERENCE: GC636-2
; CURRENT APPLICATION NUMBER: US/09/975,132A
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US 60/239,531
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: tagged IL-3 amino acid
US-09-975-132A-28

Query Match 0.8%; Score 7; DB 9; Length 173;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 KRLYARL 172
Db 5 KRLYARL 11

RESULT 18
US-09-986-480-368
; Sequence 368, Application US/09986480
; Publication No. US20030027999A1

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 143 Human Secreted Proteins
; FILE REFERENCE: PS500P1
; CURRENT APPLICATION NUMBER: US/09/986,480
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: PCT/US00/12788
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/134,068
; PRIOR FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 456
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 368
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-480-368

Query Match 0.8%; Score 7; DB 9; Length 187;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 384 PVKRELL 390
Db 31 PVKRELL 37

RESULT 19
US-10-062-254-162
; Sequence 162, Application US/10062254
; Patent No. US20020138882A1
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Edgar B
; APPLICANT: Cahoon, Rebecca E
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Fang, Yiwen
; APPLICANT: Hanke, Sabine S.
; APPLICANT: Lee, Jian-Ming
; APPLICANT: Li, Zhongsen
; APPLICANT: Miao, Guo-Hua
; APPLICANT: Morgante, Michele
; APPLICANT: Niu, Xiping
; APPLICANT: Odell, Joan
; APPLICANT: Rafalski, Antoni
; APPLICANT: Sakai, Hajime
; APPLICANT: Zheng, Peizhong
; APPLICANT: Zhu, Qun
; TITLE OF INVENTION: Polynucleotides Encoding Proteins Involved In Plant Metabolism
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/062,254
; CURRENT FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/630,346
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/146511
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/156006
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/156899
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 60/157287
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 60/169767
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 60/171054
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: 60/172958
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/171515
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: 60/173535
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 375
; SOFTWARE: Microsoft Office 97

```

; CURRENT APPLICATION NUMBER: US/09/764,855
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 140
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-855-140

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 106;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 178 NKVPRLK 184
DB 69 NKVPRLK 75

RESULT 12
US-10-042-141-91
; Sequence 91, Application US/10042141
; Publication No. US20020183503A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/10/042,141
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: 09/726,643
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 91
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-042-141-91

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 122;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 508 DRVLA1N 514
DB 45 DRVLA1N 51

RESULT 13
US-09-726-643-91
; Sequence 91, Application US/09726643
; Patent No. US20020028449A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/09/726,643
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 91
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```

US-09-726-643-91

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 122;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 508 DRVLA1N 514
DB 45 DRVLA1N 51

RESULT 14
US-09-764-868-1129
; Sequence 1129, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PAM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1129
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (125)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (138)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-764-868-1129

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 150;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 DYNFDD 91
DB 85 DYNFDD 91

RESULT 15
US-09-975-132A-26
; Sequence 26, Application US/09975132A
; Publication No. US20020182672A1
; GENERAL INFORMATION:
; APPLICANT: Kolikman, Marc
; TITLE OF INVENTION: Enhanced Secretion of a Polypeptide by a
; FILE REFERENCE: GC636-2
; CURRENT APPLICATION NUMBER: US/09/975,132A
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US 60/239,531
; PRIOR FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: IL-3 encoded by plasmid pLATITL-3
; US-09-975-132A-26

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 158;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

; Sequence 24, Application US/09746919
; Patent No. US20020013452A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Howard M.
; APPLICANT: Pontzer, Carol H.
; TITLE OF INVENTION: Interferon Tau Compositions and
; TITLE OF INVENTION: Methods of Use
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Denlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/746,919
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/045,467
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/438,753
; FILING DATE: 10-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/139,891
; FILING DATE: 19-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/847,741
; FILING DATE: 09-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/318,050
; FILING DATE: 02-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/969,890
; FILING DATE: 30-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Denlinger, Peter J.
; REGISTRATION NUMBER: 28,006
; REFERENCE/DOCKET NUMBER: 5600-0001.36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: of predicted amino acid coding sequence
; US-09-746-919-24
;
; Query Match
; Best Local Similarity 0.8%; Score 7; DB 10; Length 95;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 389 LLEQLLT 395
;      |||||
;      42 LLEQLLT 48
;
; RESULT 9
; US-09-925-297-506
; Sequence 506, Application US/09925297
; Patent No. US20020081659A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 506
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (45)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (92)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-925-297-506

```

```

; Query Match
; Best Local Similarity 0.8%; Score 7; DB 10; Length 102;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 57 PVLITP 63
;      |||||
;      65 PVLITP 71
;
; RESULT 10
; US-10-072-349-140
; Sequence 140, Application US/10072349
; Publication No. US20030054420A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA110C1
; CURRENT APPLICATION NUMBER: US/10/072,349
; PRIOR FILING DATE: 2002-02-11
; PRIOR APPLICATION removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: Patentin Ver. 3.1
; SEQ ID NO 140
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-072-349-140

```

```

; Query Match
; Best Local Similarity 0.8%; Score 7; DB 9; Length 106;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 178 NKVPRLK 184
;      |||||
;      69 NKVPRLK 75
;
; RESULT 11
; US-09-764-855-140
; Sequence 140, Application US/09764855
; Patent No. US20020119919A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA110

```

LENGTH: 41
TYPE: PRT
ORGANISM: Homo sapiens
US-09-726-643-93

Query Match 0.8%; Score 7; DB 10; Length 41;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 DRYLATN 514
|||||||
DB 4 DRYLATN 10

RESULT 4
US-10-091-504-1204
Sequence 1204, Application US/10091504
Publication No. US20030059908A1

GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091,504
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1204
LENGTH: 51
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: (20)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: misc-feature
LOCATION: (40)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-091-504-1204

Query Match 0.8%; Score 7; DB 9; Length 51;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 234 LROTALV 240
|||||||
DB 25 LROTALV 31

RESULT 5
US-09-764-869-1204
Sequence 1204, Application US/09764869
Patent No. US2002061521A1

GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007
CURRENT APPLICATION NUMBER: US/09/764,869
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1204
LENGTH: 51
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (20)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (40)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-869-1204

Query Match 0.8%; Score 7; DB 10; Length 51;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 LROTALV 240
|||||||
DB 25 LROTALV 31

RESULT 6
US-09-809-391-601

Sequence 601, Application US/09809391
Publication No. US20030049618A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002P2
CURRENT APPLICATION NUMBER: US/09/809,391
CURRENT FILING DATE: 2001-03-16
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 761
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 601
LENGTH: 69
TYPE: PRT
ORGANISM: Homo sapiens
US-09-809-391-601

Query Match 0.8%; Score 7; DB 9; Length 69;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 388 ELLEQLL 394
|||||||
DB 51 ELLEQLL 57

RESULT 7
US-09-809-391-465

Sequence 465, Application US/09809391
Publication No. US20030049618A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002P2
CURRENT APPLICATION NUMBER: US/09/809,391
CURRENT FILING DATE: 2001-03-16
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 761
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 465
LENGTH: 70
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (70)
OTHER INFORMATION: Xaa equals stop translation
US-09-809-391-465

Query Match 0.8%; Score 7; DB 9; Length 70;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 ELLEQLL 394
|||||||
DB 51 ELLEQLL 57

RESULT 8
US-09-746-919-24

```

969 6 0.7 101 9 US-09-738-626-6090 Sequence 6090, Ap
970 6 0.7 101 10 US-09-864-761-35856 Sequence 35556, A
971 6 0.7 104 9 US-09-834-794A-2 Sequence 2, Appl1
972 6 0.7 104 10 US-09-834-795A-2 Sequence 2, Appl1
973 6 0.7 105 9 US-10-023-934-89 Sequence 89, Appl1
974 6 0.7 106 9 US-10-023-896-82 Sequence 82, Appl1
975 6 0.7 107 10 US-09-864-761-43059 Sequence 43059, A
976 6 0.7 112 9 US-10-051-664-78 Sequence 78, Appl1
977 6 0.7 112 9 US-09-880-505-78 Sequence 78, Appl1
978 6 0.7 112 9 US-09-974-879-188 Sequence 188, Appl1
979 6 0.7 112 10 US-09-864-761-34124 Sequence 34124, A
980 6 0.7 113 10 US-09-925-300-1011 Sequence 1011, Ap
981 6 0.7 114 10 US-08-979-847-90 Sequence 90, Appl1
982 6 0.7 118 10 US-09-867-550-1670 Sequence 1670, Ap
983 6 0.7 119 9 US-10-101-464A-581 Sequence 581, Appl1
984 6 0.7 122 9 US-10-067-974-14 Sequence 14, Appl1
985 6 0.7 122 10 US-09-815-242-10421 Sequence 10421, A
986 6 0.7 122 10 US-09-815-242-13784 Sequence 13784, A
987 6 0.7 123 10 US-09-739-254-130 Sequence 130, App
988 6 0.7 123 10 US-09-904-615-130 Sequence 130, App
989 6 0.7 125 10 US-09-864-761-36218 Sequence 36218, A
990 6 0.7 126 9 US-09-738-626-6762 Sequence 6762, Ap
991 6 0.7 127 9 US-09-898-751A-6 Sequence 6, Appl1
992 6 0.7 127 9 US-09-738-626-5014 Sequence 5014, Ap
993 6 0.7 127 9 US-09-834-794A-1 Sequence 1, Appl1
994 6 0.7 127 9 US-10-146-496-2 Sequence 2, Appl1
995 6 0.7 127 10 US-09-813-492-2 Sequence 2, Appl1
996 6 0.7 127 10 US-09-834-795A-1 Sequence 1, Appl1
997 6 0.7 127 10 US-09-931-381A-2 Sequence 2, Appl1
998 6 0.7 130 9 US-09-925-299-1128 Sequence 1128, Ap
999 6 0.7 130 10 US-09-925-299-1128 Sequence 1128, Ap
1000 6 0.7 131 9 US-09-764-868-1156 Sequence 1156, Ap

```

ALIGNMENTS

```

RESULT 1
US-10-045-792-2
; Sequence 2, Application US/10045792
; Publication No. US20030003563A1

```

GENERAL INFORMATION:

```

APPLICANT: Vinkemeier, Dwe
Moareff, Ismail
Darnell, Jr., James E.
Kurlyan, John
TITLE OF INVENTION: A CRYSTAL OF THE N-TERMINAL DOMAIN OF A
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSER: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

```

COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/10/045,792
APPLICATION NUMBER: US/10-045,792
FILING DATE: 19-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/012,710
FILING DATE: <Unknown>

```

ATTORNEY/AGENT INFORMATION:

```

NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-194
TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 851 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-045-792-2

```

```

Query Match 1.0%; Score 9; DB 9; Length 851;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 457 ESTLEPVIE 465
Db 756 ESTLEPVIE 764

```

```

RESULT 2
US-10-042-141-93
; Sequence 93, Application US/10042141
; Publication No. US20020183503A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/10/042,141
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: 09/726,643
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 93
; LENGTH: 41
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-141-93

```

```

Query Match 0.8%; Score 7; DB 9; Length 41;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 508 DRYLVAIN 514
Db 4 DRYLVAIN 10

```

```

RESULT 3
US-09-726-643-93
; Sequence 93, Application US/09726643
; Patent No. US20020028449A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/09/726,643
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 93

```

823	7	0.8	632	9	US-10-205-894-166	Sequence 166, App	896	7	0.8	984	9	US-10-101-464A-919	Sequence 919, App
824	7	0.8	632	9	US-10-205-896-166	Sequence 166, App	897	7	0.8	1014	9	US-10-101-464A-807	Sequence 807, App
825	7	0.8	632	9	US-10-205-898-166	Sequence 166, App	898	7	0.8	1021	9	US-10-101-464A-954	Sequence 954, App
826	7	0.8	632	9	US-10-205-901-166	Sequence 166, App	899	7	0.8	1228	9	US-09-117-447-2	Sequence 2, App1
827	7	0.8	632	9	US-10-205-903-166	Sequence 166, App	900	7	0.8	2336	12	US-10-033-026-10	Sequence 10, App1
828	7	0.8	632	9	US-10-206-809-166	Sequence 166, App	901	7	0.8	2969	9	US-09-738-626-4434	Sequence 4434, App
829	7	0.8	632	9	US-10-206-910-166	Sequence 166, App	902	6	0.7	9	9	US-09-264-516A-32	Sequence 32, App1
830	7	0.8	632	9	US-10-206-911-166	Sequence 166, App	903	6	0.7	9	9	US-09-909-460-95	Sequence 95, App1
831	7	0.8	632	9	US-10-206-912-166	Sequence 166, App	904	6	0.7	9	9	US-10-119-537-1	Sequence 1, App1
832	7	0.8	632	9	US-10-206-913-166	Sequence 166, App	905	6	0.7	10	10	US-09-894-018-327	Sequence 327, App
833	7	0.8	632	9	US-10-206-914-166	Sequence 166, App	906	6	0.7	13	10	US-09-791-378-430	Sequence 430, App
834	7	0.8	632	9	US-10-206-920-166	Sequence 166, App	907	6	0.7	15	9	US-09-865-989-252	Sequence 252, App
835	7	0.8	632	9	US-10-206-921-166	Sequence 166, App	908	6	0.7	13	9	US-10-099-574A-252	Sequence 252, App
836	7	0.8	632	9	US-10-206-923-166	Sequence 166, App	909	6	0.7	15	9	US-10-099-574A-252	Sequence 252, App
837	7	0.8	632	9	US-10-206-925-166	Sequence 166, App	910	6	0.7	18	9	US-10-006-553-33	Sequence 33, App1
838	7	0.8	632	9	US-10-206-926-166	Sequence 166, App	911	6	0.7	19	9	US-09-834-794A-5	Sequence 5, App1
839	7	0.8	632	9	US-10-206-927-166	Sequence 166, App	912	6	0.7	12	10	US-09-834-795A-5	Sequence 5, App1
840	7	0.8	632	9	US-10-207-916-166	Sequence 166, App	913	6	0.7	24	9	US-09-933-999A-14	Sequence 14, App1
841	7	0.8	632	9	US-10-207-917-166	Sequence 166, App	914	6	0.7	24	10	US-10-001-883-134	Sequence 134, App
842	7	0.8	632	9	US-10-207-918-166	Sequence 166, App	915	6	0.7	25	9	US-09-864-761-22118	Sequence 284, App
843	7	0.8	632	9	US-10-207-919-166	Sequence 166, App	916	6	0.7	27	10	US-10-097-065-284	Sequence 284, App
844	7	0.8	632	9	US-10-207-920-166	Sequence 166, App	917	6	0.7	32	9	US-10-174-410-209	Sequence 209, App
845	7	0.8	632	9	US-10-207-925-166	Sequence 166, App	918	6	0.7	32	10	US-09-821-883-11	Sequence 11, App1
846	7	0.8	632	9	US-10-208-021-166	Sequence 166, App	919	6	0.7	32	10	US-10-011-585A-159	Sequence 159, App
847	7	0.8	632	9	US-10-208-022-166	Sequence 166, App	920	6	0.7	37	9	US-09-864-761-39173	Sequence 39173, App
848	7	0.8	632	9	US-10-208-023-166	Sequence 166, App	921	6	0.7	38	10	US-09-925-302-747	Sequence 747, App
849	7	0.8	632	9	US-10-208-026-166	Sequence 166, App	922	6	0.7	40	10	US-09-864-761-36789	Sequence 36789, App
850	7	0.8	632	9	US-10-208-029-166	Sequence 166, App	923	6	0.7	41	9	US-09-986-480-212	Sequence 212, App
851	7	0.8	632	9	US-10-208-030-166	Sequence 166, App	924	6	0.7	46	10	US-09-864-761-39925	Sequence 39925, App
852	7	0.8	632	9	US-10-232-232-166	Sequence 166, App	925	6	0.7	46	10	US-09-864-761-46687	Sequence 46687, App
853	7	0.8	632	9	US-09-990-427-219	Sequence 219, App	926	6	0.7	48	10	US-09-864-761-43654	Sequence 43654, App
854	7	0.8	632	9	US-09-990-439-219	Sequence 219, App	927	6	0.7	50	10	US-09-864-761-44317	Sequence 44317, App
855	7	0.8	632	9	US-10-063-538-40	Sequence 40, App1	928	6	0.7	54	9	US-10-016-157A-176	Sequence 176, App
856	7	0.8	632	9	US-10-173-693-166	Sequence 166, App	929	6	0.7	55	9	US-10-091-572-293	Sequence 293, App
857	7	0.8	632	9	US-10-174-578-166	Sequence 166, App	930	6	0.7	59	9	US-10-119-537-2	Sequence 2, App1
858	7	0.8	632	9	US-10-175-741-166	Sequence 166, App	931	6	0.7	60	9	US-10-108-605-107	Sequence 107, App
859	7	0.8	632	9	US-10-175-750-166	Sequence 166, App	932	6	0.7	62	10	US-09-864-761-41663	Sequence 41663, App
860	7	0.8	632	9	US-10-176-986-166	Sequence 166, App	933	6	0.7	66	10	US-09-925-297-689	Sequence 689, App
861	7	0.8	632	9	US-10-184-664-166	Sequence 166, App	934	6	0.7	69	10	US-09-864-761-34228	Sequence 34228, App
862	7	0.8	632	9	US-10-187-888-166	Sequence 166, App	935	6	0.7	69	10	US-09-864-761-39101	Sequence 39101, App
863	7	0.8	632	9	US-10-194-360-166	Sequence 166, App	936	6	0.7	71	9	US-10-001-835-211	Sequence 211, App
864	7	0.8	632	9	US-10-194-365-166	Sequence 166, App	937	6	0.7	71	10	US-09-864-761-37218	Sequence 37218, App
865	7	0.8	632	9	US-10-195-895-166	Sequence 166, App	938	6	0.7	71	10	US-09-864-761-42991	Sequence 42991, App
866	7	0.8	632	9	US-10-195-898-166	Sequence 166, App	939	6	0.7	71	10	US-09-864-761-48254	Sequence 48254, App
867	7	0.8	632	9	US-10-196-759-166	Sequence 166, App	940	6	0.7	73	10	US-09-864-761-33893	Sequence 33893, App
868	7	0.8	632	9	US-10-199-302-166	Sequence 166, App	941	6	0.7	75	10	US-09-779-239-4	Sequence 4, App1
869	7	0.8	632	9	US-10-201-323-166	Sequence 166, App	942	6	0.7	77	10	US-09-864-761-40148	Sequence 40148, App
870	7	0.8	632	9	US-10-205-510-166	Sequence 166, App	943	6	0.7	77	10	US-09-867-550-574	Sequence 574, App
871	7	0.8	632	9	US-10-205-891-166	Sequence 166, App	944	6	0.7	77	10	US-09-796-336A-22	Sequence 587, App
872	7	0.8	632	9	US-10-205-904-166	Sequence 166, App	945	6	0.7	78	10	US-09-864-761-37088	Sequence 37088, App
873	7	0.8	632	9	US-10-206-917-166	Sequence 166, App	946	6	0.7	79	9	US-09-335-325-11	Sequence 7, App1
874	7	0.8	632	9	US-10-207-923-166	Sequence 166, App	947	6	0.7	79	9	US-09-335-325-11	Sequence 11, App1
875	7	0.8	632	9	US-10-207-924-166	Sequence 166, App	948	6	0.7	79	9	US-10-131-241-7	Sequence 7, App1
876	7	0.8	632	9	US-10-208-028-166	Sequence 166, App	949	6	0.7	79	9	US-10-131-241-11	Sequence 11, App1
877	7	0.8	632	10	US-09-989-722-219	Sequence 219, App	950	6	0.7	79	10	US-09-761-120-7	Sequence 7, App1
878	7	0.8	632	10	US-09-989-723-219	Sequence 219, App	951	6	0.7	79	10	US-09-761-120-11	Sequence 11, App1
879	7	0.8	632	10	US-09-989-729-219	Sequence 219, App	952	6	0.7	80	10	US-09-864-761-48011	Sequence 48011, App
880	7	0.8	632	10	US-09-989-727-219	Sequence 219, App	953	6	0.7	80	10	US-09-925-297-554	Sequence 554, App
881	7	0.8	632	10	US-09-989-731-219	Sequence 219, App	954	6	0.7	81	10	US-10-080-960-20	Sequence 20, App1
882	7	0.8	632	10	US-09-989-732-219	Sequence 219, App	955	6	0.7	81	9	US-10-080-960-20	Sequence 20, App1
883	7	0.8	632	10	US-09-991-073-219	Sequence 219, App	956	6	0.7	83	9	US-09-759-130B-60	Sequence 60, App1
884	7	0.8	632	10	US-09-991-163-219	Sequence 219, App	957	6	0.7	84	10	US-09-925-300-1336	Sequence 1336, App
885	7	0.8	632	10	US-09-991-163-219	Sequence 219, App	958	6	0.7	88	10	US-09-867-550-744	Sequence 744, App
886	7	0.8	632	10	US-09-993-604-219	Sequence 219, App	959	6	0.7	88	10	US-09-864-761-43431	Sequence 43431, App
887	7	0.8	632	10	US-09-990-456-219	Sequence 219, App	960	6	0.7	90	10	US-09-864-761-46509	Sequence 46509, App
888	7	0.8	632	10	US-09-989-721-219	Sequence 219, App	961	6	0.7	92	9	US-10-125-258-78	Sequence 78, App1
889	7	0.8	632	12	US-10-006-867-40	Sequence 40, App1	962	6	0.7	92	9	US-10-125-258-80	Sequence 80, App1
890	7	0.8	632	12	US-10-052-586-166	Sequence 166, App	963	6	0.7	95	9	US-09-877-843-27	Sequence 27, App1
891	7	0.8	655	9	US-10-101-464A-70	Sequence 70, App1	964	6	0.7	96	9	US-09-877-843-25	Sequence 25, App1
892	7	0.8	696	9	US-09-764-868-959	Sequence 959, App	965	6	0.7	97	9	US-09-877-843-26	Sequence 26, App1
893	7	0.8	749	10	US-09-899-569A-2	Sequence 2, App1	966	6	0.7	98	9	US-09-820-843A-55	Sequence 55, App1
894	7	0.8	819	10	US-09-833-790-425	Sequence 425, App	967	6	0.7	98	9	US-10-101-464A-654	Sequence 654, App
895	7	0.8	836	10	US-09-899-569A-4	Sequence 4, App1	968	6	0.7	99	9	US-10-083-357-827	Sequence 827, App

677	7	0.8	632	9	US-10-199-305-166	Sequence 166, App	750	7	0.8	632	9	US-10-205-893-166	Sequence 166, App
678	7	0.8	632	9	US-10-199-306-166	Sequence 166, App	751	7	0.8	632	9	US-10-205-897-166	Sequence 166, App
679	7	0.8	632	9	US-10-199-310-166	Sequence 166, App	752	7	0.8	632	9	US-09-993-469-219	Sequence 219, App
680	7	0.8	632	9	US-10-199-311-166	Sequence 166, App	753	7	0.8	632	9	US-09-993-448-219	Sequence 219, App
681	7	0.8	632	9	US-10-199-314-166	Sequence 166, App	754	7	0.8	632	9	US-09-997-542-219	Sequence 219, App
682	7	0.8	632	9	US-10-199-317-166	Sequence 166, App	755	7	0.8	632	9	US-10-063-567-40	Sequence 40, App
683	7	0.8	632	9	US-10-199-665-166	Sequence 166, App	756	7	0.8	632	9	US-10-174-571-166	Sequence 166, App
684	7	0.8	632	9	US-10-199-666-166	Sequence 166, App	757	7	0.8	632	9	US-10-174-746-166	Sequence 166, App
685	7	0.8	632	9	US-10-199-669-166	Sequence 166, App	758	7	0.8	632	9	US-10-176-923-166	Sequence 166, App
686	7	0.8	632	9	US-10-201-534-166	Sequence 166, App	759	7	0.8	632	9	US-10-183-011-166	Sequence 166, App
687	7	0.8	632	9	US-10-201-770-166	Sequence 166, App	760	7	0.8	632	9	US-10-187-742-166	Sequence 166, App
688	7	0.8	632	9	US-10-201-855-166	Sequence 166, App	761	7	0.8	632	9	US-10-184-639-166	Sequence 166, App
689	7	0.8	632	9	US-10-201-856-166	Sequence 166, App	762	7	0.8	632	9	US-10-184-633-166	Sequence 166, App
690	7	0.8	632	9	US-10-202-469-166	Sequence 166, App	763	7	0.8	632	9	US-10-187-748-166	Sequence 166, App
691	7	0.8	632	9	US-10-202-470-166	Sequence 166, App	764	7	0.8	632	9	US-10-188-766-166	Sequence 166, App
692	7	0.8	632	9	US-10-202-476-166	Sequence 166, App	765	7	0.8	632	9	US-10-188-771-166	Sequence 166, App
693	7	0.8	632	9	US-10-202-934-166	Sequence 166, App	766	7	0.8	632	9	US-10-192-008-166	Sequence 166, App
694	7	0.8	632	9	US-10-202-935-166	Sequence 166, App	767	7	0.8	632	9	US-10-192-009-166	Sequence 166, App
695	7	0.8	632	9	US-10-202-936-166	Sequence 166, App	768	7	0.8	632	9	US-10-192-012-166	Sequence 166, App
696	7	0.8	632	9	US-10-202-939-166	Sequence 166, App	769	7	0.8	632	9	US-10-192-014-166	Sequence 166, App
697	7	0.8	632	9	US-10-205-504-166	Sequence 166, App	770	7	0.8	632	9	US-10-192-016-166	Sequence 166, App
698	7	0.8	632	9	US-10-205-509-166	Sequence 166, App	771	7	0.8	632	9	US-10-194-428-166	Sequence 166, App
699	7	0.8	632	9	US-10-205-895-166	Sequence 166, App	772	7	0.8	632	9	US-10-194-459-166	Sequence 166, App
700	7	0.8	632	9	US-10-205-899-166	Sequence 166, App	773	7	0.8	632	9	US-10-194-458-166	Sequence 166, App
701	7	0.8	632	9	US-10-205-900-166	Sequence 166, App	774	7	0.8	632	9	US-10-194-481-166	Sequence 166, App
702	7	0.8	632	9	US-10-205-909-166	Sequence 166, App	775	7	0.8	632	9	US-10-194-395-166	Sequence 166, App
703	7	0.8	632	9	US-10-227-693-40	Sequence 40, App1	776	7	0.8	632	9	US-10-194-424-166	Sequence 166, App
704	7	0.8	632	9	US-09-990-443-219	Sequence 219, App	777	7	0.8	632	9	US-10-194-458-166	Sequence 166, App
705	7	0.8	632	9	US-09-990-726-219	Sequence 219, App	778	7	0.8	632	9	US-10-194-459-166	Sequence 166, App
706	7	0.8	632	9	US-09-997-559-219	Sequence 219, App	779	7	0.8	632	9	US-10-195-886-166	Sequence 166, App
707	7	0.8	632	9	US-09-997-601-219	Sequence 219, App	780	7	0.8	632	9	US-10-195-886-166	Sequence 166, App
708	7	0.8	632	9	US-10-183-002-166	Sequence 219, App	781	7	0.8	632	9	US-10-196-741-166	Sequence 166, App
709	7	0.8	632	9	US-10-183-002-166	Sequence 166, App	782	7	0.8	632	9	US-10-196-753-166	Sequence 166, App
710	7	0.8	632	9	US-10-184-621-166	Sequence 166, App	783	7	0.8	632	9	US-10-196-753-166	Sequence 166, App
711	7	0.8	632	9	US-10-184-638-166	Sequence 166, App	784	7	0.8	632	9	US-10-196-753-166	Sequence 166, App
712	7	0.8	632	9	US-10-187-887-166	Sequence 166, App	785	7	0.8	632	9	US-10-196-751-166	Sequence 166, App
713	7	0.8	632	9	US-10-194-461-166	Sequence 166, App	786	7	0.8	632	9	US-10-197-698-166	Sequence 166, App
714	7	0.8	632	9	US-10-195-892-166	Sequence 166, App	787	7	0.8	632	9	US-10-197-711-166	Sequence 166, App
715	7	0.8	632	9	US-10-196-751-166	Sequence 166, App	788	7	0.8	632	9	US-10-197-696-166	Sequence 166, App
716	7	0.8	632	9	US-10-197-694-166	Sequence 166, App	789	7	0.8	632	9	US-10-197-703-166	Sequence 166, App
717	7	0.8	632	9	US-10-197-697-166	Sequence 166, App	790	7	0.8	632	9	US-10-197-703-166	Sequence 166, App
718	7	0.8	632	9	US-10-197-707-166	Sequence 166, App	791	7	0.8	632	9	US-10-198-761-166	Sequence 166, App
719	7	0.8	632	9	US-10-199-303-166	Sequence 166, App	792	7	0.8	632	9	US-10-198-761-166	Sequence 166, App
720	7	0.8	632	9	US-10-199-318-166	Sequence 166, App	793	7	0.8	632	9	US-10-198-761-166	Sequence 166, App
721	7	0.8	632	9	US-10-199-458-166	Sequence 166, App	794	7	0.8	632	9	US-10-198-761-166	Sequence 166, App
722	7	0.8	632	9	US-10-199-462-166	Sequence 166, App	795	7	0.8	632	9	US-10-198-762-166	Sequence 166, App
723	7	0.8	632	9	US-10-201-324-166	Sequence 166, App	796	7	0.8	632	9	US-10-198-767-166	Sequence 166, App
724	7	0.8	632	9	US-10-201-324-166	Sequence 166, App	797	7	0.8	632	9	US-10-199-301-166	Sequence 166, App
725	7	0.8	632	9	US-10-201-528-166	Sequence 166, App	798	7	0.8	632	9	US-10-199-307-166	Sequence 166, App
726	7	0.8	632	9	US-10-201-529-166	Sequence 166, App	799	7	0.8	632	9	US-10-199-312-166	Sequence 166, App
727	7	0.8	632	9	US-10-201-530-166	Sequence 166, App	800	7	0.8	632	9	US-10-199-315-166	Sequence 166, App
728	7	0.8	632	9	US-10-202-408-166	Sequence 166, App	801	7	0.8	632	9	US-10-199-417-166	Sequence 166, App
729	7	0.8	632	9	US-10-202-408-166	Sequence 166, App	802	7	0.8	632	9	US-10-199-457-166	Sequence 166, App
730	7	0.8	632	9	US-10-202-408-166	Sequence 166, App	803	7	0.8	632	9	US-10-201-321-166	Sequence 166, App
731	7	0.8	632	9	US-10-202-411-166	Sequence 166, App	804	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
732	7	0.8	632	9	US-10-202-411-166	Sequence 166, App	805	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
733	7	0.8	632	9	US-10-202-472-166	Sequence 166, App	806	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
734	7	0.8	632	9	US-10-205-502-166	Sequence 166, App	807	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
735	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	808	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
736	7	0.8	632	9	US-10-205-511-166	Sequence 166, App	809	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
737	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	810	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
738	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	811	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
739	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	812	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
740	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	813	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
741	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	814	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
742	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	815	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
743	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	816	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
744	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	817	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
745	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	818	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
746	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	819	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
747	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	820	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
748	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	821	7	0.8	632	9	US-10-201-322-166	Sequence 166, App
749	7	0.8	632	9	US-10-205-507-166	Sequence 166, App	822	7	0.8	632	9	US-10-205-512-166	Sequence 166, App
750	7	0.8	632	9	US-10-205-512-166	Sequence 166, App							

531	7	0.8	632	9	US-10-180-555-166	Sequence 166, App	604	7	0.8	632	9	US-10-179-516-166	Sequence 166, App
532	7	0.8	632	9	US-10-180-559-166	Sequence 166, App	605	7	0.8	632	9	US-10-179-525-166	Sequence 166, App
533	7	0.8	632	9	US-10-181-000-166	Sequence 166, App	606	7	0.8	632	9	US-10-179-535-166	Sequence 166, App
534	7	0.8	632	9	US-10-183-010-166	Sequence 166, App	607	7	0.8	632	9	US-10-180-540-166	Sequence 166, App
535	7	0.8	632	9	US-10-183-012-166	Sequence 166, App	608	7	0.8	632	9	US-10-180-545-166	Sequence 166, App
536	7	0.8	632	9	US-10-184-614-166	Sequence 166, App	609	7	0.8	632	9	US-10-183-006-166	Sequence 166, App
537	7	0.8	632	9	US-10-184-623-166	Sequence 166, App	610	7	0.8	632	9	US-10-183-008-166	Sequence 166, App
538	7	0.8	632	9	US-10-184-635-166	Sequence 166, App	611	7	0.8	632	9	US-10-183-017-166	Sequence 166, App
539	7	0.8	632	9	US-10-184-637-166	Sequence 166, App	612	7	0.8	632	9	US-10-183-019-166	Sequence 166, App
540	7	0.8	632	9	US-10-184-646-166	Sequence 166, App	613	7	0.8	632	9	US-10-184-618-166	Sequence 166, App
541	7	0.8	632	9	US-10-184-647-166	Sequence 166, App	614	7	0.8	632	9	US-10-184-623-166	Sequence 166, App
542	7	0.8	632	9	US-10-184-652-166	Sequence 166, App	615	7	0.8	632	9	US-10-184-626-166	Sequence 166, App
543	7	0.8	632	9	US-10-187-594-166	Sequence 166, App	616	7	0.8	632	9	US-10-184-627-166	Sequence 166, App
544	7	0.8	632	9	US-10-187-596-166	Sequence 166, App	617	7	0.8	632	9	US-10-184-645-166	Sequence 166, App
545	7	0.8	632	9	US-10-187-745-166	Sequence 166, App	618	7	0.8	632	9	US-10-184-654-166	Sequence 166, App
546	7	0.8	632	9	US-10-187-885-166	Sequence 166, App	619	7	0.8	632	9	US-10-184-655-166	Sequence 166, App
547	7	0.8	632	9	US-10-187-886-166	Sequence 166, App	620	7	0.8	632	9	US-10-188-774-166	Sequence 166, App
548	7	0.8	632	9	US-10-199-464-166	Sequence 166, App	621	7	0.8	632	9	US-10-188-775-166	Sequence 166, App
549	7	0.8	632	9	US-10-176-751-166	Sequence 166, App	622	7	0.8	632	9	US-10-194-462-166	Sequence 166, App
550	7	0.8	632	9	US-10-176-760-166	Sequence 166, App	623	7	0.8	632	9	US-10-195-902-166	Sequence 166, App
551	7	0.8	632	9	US-10-176-990-166	Sequence 166, App	624	7	0.8	632	9	US-10-196-743-166	Sequence 166, App
552	7	0.8	632	9	US-10-180-541-166	Sequence 166, App	625	7	0.8	632	9	US-10-196-745-166	Sequence 166, App
553	7	0.8	632	9	US-10-180-542-166	Sequence 166, App	626	7	0.8	632	9	US-10-196-760-166	Sequence 166, App
554	7	0.8	632	9	US-10-180-548-166	Sequence 166, App	627	7	0.8	632	9	US-10-196-762-166	Sequence 166, App
555	7	0.8	632	9	US-10-180-551-166	Sequence 166, App	628	7	0.8	632	9	US-10-197-693-166	Sequence 219, App
556	7	0.8	632	9	US-10-180-998-166	Sequence 166, App	629	7	0.8	632	9	US-09-998-156-219	Sequence 219, App
557	7	0.8	632	9	US-10-180-999-166	Sequence 166, App	630	7	0.8	632	9	US-09-998-156-219	Sequence 219, App
558	7	0.8	632	9	US-10-183-013-166	Sequence 166, App	631	7	0.8	632	9	US-10-176-484-166	Sequence 166, App
559	7	0.8	632	9	US-10-184-612-166	Sequence 166, App	632	7	0.8	632	9	US-10-176-917-166	Sequence 166, App
560	7	0.8	632	9	US-10-184-616-166	Sequence 166, App	633	7	0.8	632	9	US-10-176-982-166	Sequence 166, App
561	7	0.8	632	9	US-10-184-617-166	Sequence 166, App	634	7	0.8	632	9	US-10-179-506-166	Sequence 166, App
562	7	0.8	632	9	US-10-184-622-166	Sequence 166, App	635	7	0.8	632	9	US-10-179-513-166	Sequence 166, App
563	7	0.8	632	9	US-10-184-629-166	Sequence 166, App	636	7	0.8	632	9	US-10-179-514-166	Sequence 166, App
564	7	0.8	632	9	US-10-184-630-166	Sequence 166, App	637	7	0.8	632	9	US-10-179-514-166	Sequence 166, App
565	7	0.8	632	9	US-10-184-631-166	Sequence 166, App	638	7	0.8	632	9	US-10-179-522-166	Sequence 166, App
566	7	0.8	632	9	US-10-184-636-166	Sequence 166, App	639	7	0.8	632	9	US-10-180-556-166	Sequence 166, App
567	7	0.8	632	9	US-10-184-632-166	Sequence 166, App	640	7	0.8	632	9	US-10-180-560-166	Sequence 166, App
568	7	0.8	632	9	US-10-184-636-166	Sequence 166, App	641	7	0.8	632	9	US-10-183-015-166	Sequence 166, App
569	7	0.8	632	9	US-10-184-640-166	Sequence 166, App	642	7	0.8	632	9	US-10-184-615-166	Sequence 166, App
570	7	0.8	632	9	US-10-184-650-166	Sequence 166, App	643	7	0.8	632	9	US-10-184-620-166	Sequence 166, App
571	7	0.8	632	9	US-10-184-651-166	Sequence 166, App	644	7	0.8	632	9	US-10-184-643-166	Sequence 166, App
572	7	0.8	632	9	US-10-187-588-166	Sequence 166, App	645	7	0.8	632	9	US-10-184-656-166	Sequence 166, App
573	7	0.8	632	9	US-10-187-597-166	Sequence 166, App	646	7	0.8	632	9	US-10-192-010-166	Sequence 166, App
574	7	0.8	632	9	US-10-187-598-166	Sequence 166, App	647	7	0.8	632	9	US-10-195-894-166	Sequence 166, App
575	7	0.8	632	9	US-10-187-600-166	Sequence 166, App	648	7	0.8	632	9	US-10-205-808-166	Sequence 166, App
576	7	0.8	632	9	US-10-187-601-166	Sequence 166, App	649	7	0.8	632	9	US-09-991-157-219	Sequence 219, App
577	7	0.8	632	9	US-10-187-602-166	Sequence 166, App	650	7	0.8	632	9	US-09-991-172-219	Sequence 219, App
578	7	0.8	632	9	US-10-187-603-166	Sequence 166, App	651	7	0.8	632	9	US-09-997-514-219	Sequence 219, App
579	7	0.8	632	9	US-10-187-741-166	Sequence 166, App	652	7	0.8	632	9	US-09-997-514-219	Sequence 219, App
580	7	0.8	632	9	US-10-187-743-166	Sequence 166, App	653	7	0.8	632	9	US-09-997-514-219	Sequence 219, App
581	7	0.8	632	9	US-10-187-746-166	Sequence 166, App	654	7	0.8	632	9	US-10-063-518-40	Sequence 40, App1
582	7	0.8	632	9	US-10-187-747-166	Sequence 166, App	655	7	0.8	632	9	US-10-063-518-40	Sequence 40, App1
583	7	0.8	632	9	US-10-187-751-166	Sequence 166, App	656	7	0.8	632	9	US-10-184-619-166	Sequence 166, App
584	7	0.8	632	9	US-10-187-753-166	Sequence 166, App	657	7	0.8	632	9	US-10-186-619-166	Sequence 166, App
585	7	0.8	632	9	US-10-187-754-166	Sequence 166, App	658	7	0.8	632	9	US-10-186-620-166	Sequence 166, App
586	7	0.8	632	9	US-10-187-757-166	Sequence 166, App	659	7	0.8	632	9	US-10-186-621-166	Sequence 166, App
587	7	0.8	632	9	US-10-187-884-166	Sequence 166, App	660	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
588	7	0.8	632	9	US-10-188-767-166	Sequence 166, App	661	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
589	7	0.8	632	9	US-10-188-769-166	Sequence 166, App	662	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
590	7	0.8	632	9	US-10-188-770-166	Sequence 166, App	663	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
591	7	0.8	632	9	US-10-188-773-166	Sequence 166, App	664	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
592	7	0.8	632	9	US-10-188-781-166	Sequence 166, App	665	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
593	7	0.8	632	9	US-10-194-351-166	Sequence 166, App	666	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
594	7	0.8	632	9	US-10-194-423-166	Sequence 166, App	667	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
595	7	0.8	632	9	US-10-195-897-166	Sequence 166, App	668	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
596	7	0.8	632	9	US-10-195-901-166	Sequence 166, App	669	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
597	7	0.8	632	9	US-10-196-756-166	Sequence 166, App	670	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
598	7	0.8	632	9	US-09-989-726-219	Sequence 219, App	671	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
599	7	0.8	632	9	US-10-173-708-166	Sequence 166, App	672	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
600	7	0.8	632	9	US-10-176-478-166	Sequence 166, App	673	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
601	7	0.8	632	9	US-10-176-748-166	Sequence 166, App	674	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
602	7	0.8	632	9	US-10-176-916-166	Sequence 166, App	675	7	0.8	632	9	US-10-187-750-166	Sequence 166, App
603	7	0.8	632	9	US-10-179-507-166	Sequence 166, App	676	7	0.8	632	9	US-10-187-750-166	Sequence 166, App

385	7	0.8	261	9	US-10-173-693-356	Sequence 356, App	458	7	0.8	516	10	US-09-883-797-14	Sequence 14, Appl
386	7	0.8	261	9	US-10-174-578-356	Sequence 356, App	459	7	0.8	534	9	US-09-975-719-67	Sequence 67, Appl
387	7	0.8	261	9	US-10-175-741-356	Sequence 356, App	460	7	0.8	599	9	US-09-738-626-5182	Sequence 5182, Ap
388	7	0.8	261	9	US-10-175-750-356	Sequence 356, App	461	7	0.8	632	9	US-09-992-598-219	Sequence 219, App
389	7	0.8	261	9	US-10-176-986-356	Sequence 356, App	462	7	0.8	632	9	US-09-989-293A-219	Sequence 219, App
390	7	0.8	261	9	US-10-184-641-356	Sequence 356, App	463	7	0.8	632	9	US-10-063-547-40	Sequence 40, Appl
391	7	0.8	261	9	US-10-187-888-356	Sequence 356, App	464	7	0.8	632	9	US-09-989-735-219	Sequence 219, App
392	7	0.8	261	9	US-10-194-360-356	Sequence 356, App	465	7	0.8	632	9	US-09-990-44A-219	Sequence 219, App
393	7	0.8	261	9	US-10-194-365-356	Sequence 356, App	466	7	0.8	632	9	US-09-989-730-219	Sequence 219, App
394	7	0.8	261	9	US-10-195-895-356	Sequence 356, App	467	7	0.8	632	9	US-09-990-436-219	Sequence 219, App
395	7	0.8	261	9	US-10-195-898-356	Sequence 356, App	468	7	0.8	632	9	US-09-991-181-219	Sequence 219, App
396	7	0.8	261	9	US-10-196-759-356	Sequence 356, App	469	7	0.8	632	9	US-09-993-687-219	Sequence 219, App
397	7	0.8	261	9	US-10-199-302-356	Sequence 356, App	470	7	0.8	632	9	US-09-989-734-219	Sequence 219, App
398	7	0.8	261	9	US-10-201-323-356	Sequence 356, App	471	7	0.8	632	9	US-09-997-653-219	Sequence 219, App
399	7	0.8	261	9	US-10-205-510-356	Sequence 356, App	472	7	0.8	632	9	US-10-174-599-166	Sequence 166, App
400	7	0.8	261	9	US-10-203-891-356	Sequence 356, App	473	7	0.8	632	9	US-10-176-738-166	Sequence 166, App
401	7	0.8	261	9	US-10-205-904-356	Sequence 356, App	474	7	0.8	632	9	US-10-063-616-40	Sequence 40, Appl
402	7	0.8	261	9	US-10-206-917-356	Sequence 356, App	475	7	0.8	632	9	US-10-173-706-166	Sequence 166, App
403	7	0.8	261	9	US-10-207-923-356	Sequence 356, App	476	7	0.8	632	9	US-10-175-737-166	Sequence 166, App
404	7	0.8	261	9	US-10-207-924-356	Sequence 356, App	477	7	0.8	632	9	US-10-063-502-40	Sequence 40, Appl
405	7	0.8	261	9	US-10-208-028-356	Sequence 356, App	478	7	0.8	632	9	US-10-173-706-166	Sequence 166, App
406	7	0.8	261	12	US-10-006-867-118	Sequence 118, App	479	7	0.8	632	9	US-10-175-738-166	Sequence 166, App
407	7	0.8	261	12	US-10-052-586-356	Sequence 356, App	480	7	0.8	632	9	US-10-176-482-166	Sequence 166, App
408	7	0.8	261	9	US-09-981-353-90	Sequence 90, Appl	481	7	0.8	632	9	US-10-176-757-166	Sequence 166, App
409	7	0.8	308	9	US-09-738-626-6317	Sequence 509, App	482	7	0.8	632	9	US-10-175-752-166	Sequence 166, App
410	7	0.8	313	10	US-09-886-055-509	Sequence 10321, A	483	7	0.8	632	9	US-10-176-913-166	Sequence 166, App
411	7	0.8	323	10	US-09-815-242-10321	Sequence 10321, A	484	7	0.8	632	9	US-10-180-552-166	Sequence 166, App
412	7	0.8	323	10	US-09-815-242-13734	Sequence 13734, A	485	7	0.8	632	9	US-10-180-557-166	Sequence 166, App
413	7	0.8	343	9	US-10-028-072-162	Sequence 162, App	486	7	0.8	632	9	US-09-990-438-219	Sequence 219, App
414	7	0.8	343	9	US-10-121-049-162	Sequence 162, App	487	7	0.8	632	9	US-09-997-428-219	Sequence 219, App
415	7	0.8	343	9	US-10-123-904-162	Sequence 162, App	488	7	0.8	632	9	US-09-997-666-219	Sequence 219, App
416	7	0.8	343	9	US-10-140-470-162	Sequence 162, App	489	7	0.8	632	9	US-10-173-780-166	Sequence 166, App
417	7	0.8	343	9	US-10-175-746-162	Sequence 162, App	490	7	0.8	632	9	US-10-174-572-166	Sequence 166, App
418	7	0.8	343	9	US-10-176-918-162	Sequence 162, App	491	7	0.8	632	9	US-10-174-572-166	Sequence 166, App
419	7	0.8	343	9	US-10-176-921-162	Sequence 162, App	492	7	0.8	632	9	US-10-174-579-166	Sequence 166, App
420	7	0.8	343	9	US-10-137-865-162	Sequence 162, App	493	7	0.8	632	9	US-10-174-588-166	Sequence 166, App
421	7	0.8	343	9	US-10-140-474-162	Sequence 162, App	494	7	0.8	632	9	US-10-175-739-166	Sequence 166, App
422	7	0.8	343	9	US-10-143-431-162	Sequence 162, App	495	7	0.8	632	9	US-10-175-740-166	Sequence 166, App
423	7	0.8	343	9	US-10-140-002-162	Sequence 162, App	496	7	0.8	632	9	US-10-175-743-166	Sequence 166, App
424	7	0.8	343	9	US-10-142-419-162	Sequence 162, App	497	7	0.8	632	9	US-10-176-488-166	Sequence 166, App
425	7	0.8	343	9	US-10-123-262-162	Sequence 162, App	498	7	0.8	632	9	US-10-176-492-166	Sequence 166, App
426	7	0.8	343	9	US-10-121-050-162	Sequence 162, App	499	7	0.8	632	9	US-10-176-747-166	Sequence 166, App
427	7	0.8	343	9	US-10-141-755-162	Sequence 162, App	500	7	0.8	632	9	US-10-176-985-166	Sequence 166, App
428	7	0.8	343	9	US-10-143-032-162	Sequence 162, App	501	7	0.8	632	9	US-10-176-987-166	Sequence 166, App
429	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	502	7	0.8	632	9	US-10-176-991-166	Sequence 166, App
430	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	503	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
431	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	504	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
432	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	505	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
433	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	506	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
434	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	507	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
435	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	508	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
436	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	509	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
437	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	510	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
438	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	511	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
439	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	512	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
440	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	513	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
441	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	514	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
442	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	515	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
443	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	516	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
444	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	517	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
445	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	518	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
446	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	519	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
447	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	520	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
448	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	521	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
449	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	522	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
450	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	523	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
451	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	524	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
452	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	525	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
453	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	526	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
454	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	527	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
455	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	528	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
456	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	529	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
457	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	530	7	0.8	632	9	US-10-176-992-166	Sequence 166, App

93	7	0.8	261	9	US-10-187-886-356	Sequence 356, App
94	7	0.8	261	9	US-10-199-464-356	Sequence 356, App
95	7	0.8	261	9	US-10-176-751-356	Sequence 356, App
96	7	0.8	261	9	US-10-176-760-356	Sequence 356, App
97	7	0.8	261	9	US-10-176-990-356	Sequence 356, App
98	7	0.8	261	9	US-10-180-541-356	Sequence 356, App
99	7	0.8	261	9	US-10-180-542-356	Sequence 356, App
100	7	0.8	261	9	US-10-180-548-356	Sequence 356, App
101	7	0.8	261	9	US-10-180-551-356	Sequence 356, App
102	7	0.8	261	9	US-10-180-998-356	Sequence 356, App
103	7	0.8	261	9	US-10-180-999-356	Sequence 356, App
104	7	0.8	261	9	US-10-183-013-356	Sequence 356, App
105	7	0.8	261	9	US-10-184-612-356	Sequence 356, App
106	7	0.8	261	9	US-10-184-615-356	Sequence 356, App
107	7	0.8	261	9	US-10-184-617-356	Sequence 356, App
108	7	0.8	261	9	US-10-184-622-356	Sequence 356, App
109	7	0.8	261	9	US-10-184-628-356	Sequence 356, App
110	7	0.8	261	9	US-10-184-630-356	Sequence 356, App
111	7	0.8	261	9	US-10-184-631-356	Sequence 356, App
112	7	0.8	261	9	US-10-184-632-356	Sequence 356, App
113	7	0.8	261	9	US-10-184-636-356	Sequence 356, App
114	7	0.8	261	9	US-10-184-640-356	Sequence 356, App
115	7	0.8	261	9	US-10-184-650-356	Sequence 356, App
116	7	0.8	261	9	US-10-184-651-356	Sequence 356, App
117	7	0.8	261	9	US-10-187-597-356	Sequence 356, App
118	7	0.8	261	9	US-10-187-598-356	Sequence 356, App
119	7	0.8	261	9	US-10-187-601-356	Sequence 356, App
120	7	0.8	261	9	US-10-187-602-356	Sequence 356, App
121	7	0.8	261	9	US-10-187-598-356	Sequence 356, App
122	7	0.8	261	9	US-10-187-601-356	Sequence 356, App
123	7	0.8	261	9	US-10-187-602-356	Sequence 356, App
124	7	0.8	261	9	US-10-187-603-356	Sequence 356, App
125	7	0.8	261	9	US-10-187-741-356	Sequence 356, App
126	7	0.8	261	9	US-10-187-743-356	Sequence 356, App
127	7	0.8	261	9	US-10-187-746-356	Sequence 356, App
128	7	0.8	261	9	US-10-187-747-356	Sequence 356, App
129	7	0.8	261	9	US-10-187-751-356	Sequence 356, App
130	7	0.8	261	9	US-10-187-753-356	Sequence 356, App
131	7	0.8	261	9	US-10-187-753-356	Sequence 356, App
132	7	0.8	261	9	US-10-187-754-356	Sequence 356, App
133	7	0.8	261	9	US-10-187-757-356	Sequence 356, App
134	7	0.8	261	9	US-10-187-884-356	Sequence 356, App
135	7	0.8	261	9	US-10-188-767-356	Sequence 356, App
136	7	0.8	261	9	US-10-188-769-356	Sequence 356, App
137	7	0.8	261	9	US-10-188-770-356	Sequence 356, App
138	7	0.8	261	9	US-10-188-773-356	Sequence 356, App
139	7	0.8	261	9	US-10-188-781-356	Sequence 356, App
140	7	0.8	261	9	US-10-194-361-356	Sequence 356, App
141	7	0.8	261	9	US-10-194-423-356	Sequence 356, App
142	7	0.8	261	9	US-10-195-897-356	Sequence 356, App
143	7	0.8	261	9	US-10-195-901-356	Sequence 356, App
144	7	0.8	261	9	US-10-196-756-356	Sequence 356, App
145	7	0.8	261	9	US-10-196-756-356	Sequence 356, App
146	7	0.8	261	9	US-10-173-708-356	Sequence 356, App
147	7	0.8	261	9	US-10-176-479-356	Sequence 356, App
148	7	0.8	261	9	US-10-176-748-356	Sequence 356, App
149	7	0.8	261	9	US-10-176-916-356	Sequence 356, App
150	7	0.8	261	9	US-10-179-507-356	Sequence 356, App
151	7	0.8	261	9	US-10-179-516-356	Sequence 356, App
152	7	0.8	261	9	US-10-179-519-356	Sequence 356, App
153	7	0.8	261	9	US-10-183-008-356	Sequence 356, App
154	7	0.8	261	9	US-10-183-017-356	Sequence 356, App
155	7	0.8	261	9	US-10-183-008-356	Sequence 356, App
156	7	0.8	261	9	US-10-183-017-356	Sequence 356, App
157	7	0.8	261	9	US-10-183-019-356	Sequence 356, App
158	7	0.8	261	9	US-10-184-618-356	Sequence 356, App
159	7	0.8	261	9	US-10-184-619-356	Sequence 356, App
160	7	0.8	261	9	US-10-184-625-356	Sequence 356, App
161	7	0.8	261	9	US-10-184-626-356	Sequence 356, App
162	7	0.8	261	9	US-10-184-627-356	Sequence 356, App
163	7	0.8	261	9	US-10-184-645-356	Sequence 356, App
164	7	0.8	261	9	US-10-184-654-356	Sequence 356, App
165	7	0.8	261	9	US-10-184-655-356	Sequence 356, App
	7	0.8	261	9	US-10-188-774-356	Sequence 356, App

166	7	0.8	261	9	US-10-188-775-356	Sequence 356, App
167	7	0.8	261	9	US-10-194-462-356	Sequence 356, App
168	7	0.8	261	9	US-10-195-902-356	Sequence 356, App
169	7	0.8	261	9	US-10-196-743-356	Sequence 356, App
170	7	0.8	261	9	US-10-196-745-356	Sequence 356, App
171	7	0.8	261	9	US-10-196-760-356	Sequence 356, App
172	7	0.8	261	9	US-10-196-762-356	Sequence 356, App
173	7	0.8	261	9	US-10-197-695-356	Sequence 356, App
174	7	0.8	261	9	US-10-006-856A-356	Sequence 356, App
175	7	0.8	261	9	US-10-176-484-356	Sequence 356, App
176	7	0.8	261	9	US-10-176-753-356	Sequence 356, App
177	7	0.8	261	9	US-10-176-917-356	Sequence 356, App
178	7	0.8	261	9	US-10-176-982-356	Sequence 356, App
179	7	0.8	261	9	US-10-179-506-356	Sequence 356, App
180	7	0.8	261	9	US-10-179-513-356	Sequence 356, App
181	7	0.8	261	9	US-10-184-615-356	Sequence 356, App
182	7	0.8	261	9	US-10-184-643-356	Sequence 356, App
183	7	0.8	261	9	US-10-180-556-356	Sequence 356, App
184	7	0.8	261	9	US-10-179-522-356	Sequence 356, App
185	7	0.8	261	9	US-10-183-015-356	Sequence 356, App
186	7	0.8	261	9	US-10-184-615-356	Sequence 356, App
187	7	0.8	261	9	US-10-184-620-356	Sequence 356, App
188	7	0.8	261	9	US-10-184-620-356	Sequence 356, App
189	7	0.8	261	9	US-10-184-653-356	Sequence 356, App
190	7	0.8	261	9	US-10-184-656-356	Sequence 356, App
191	7	0.8	261	9	US-10-192-010-356	Sequence 356, App
192	7	0.8	261	9	US-10-192-894-356	Sequence 356, App
193	7	0.8	261	9	US-10-205-908-356	Sequence 356, App
194	7	0.8	261	9	US-10-063-518-118	Sequence 118, App
195	7	0.8	261	9	US-10-063-598-118	Sequence 118, App
196	7	0.8	261	9	US-10-184-619-356	Sequence 356, App
197	7	0.8	261	9	US-10-186-855-356	Sequence 356, App
198	7	0.8	261	9	US-10-187-599-356	Sequence 356, App
199	7	0.8	261	9	US-10-187-750-356	Sequence 356, App
200	7	0.8	261	9	US-10-188-780-356	Sequence 356, App
201	7	0.8	261	9	US-10-192-015-356	Sequence 356, App
202	7	0.8	261	9	US-10-194-425-356	Sequence 356, App
203	7	0.8	261	9	US-10-194-485-356	Sequence 356, App
204	7	0.8	261	9	US-10-195-885-356	Sequence 356, App
205	7	0.8	261	9	US-10-195-885-356	Sequence 356, App
206	7	0.8	261	9	US-10-195-899-356	Sequence 356, App
207	7	0.8	261	9	US-10-196-748-356	Sequence 356, App
208	7	0.8	261	9	US-10-196-750-356	Sequence 356, App
209	7	0.8	261	9	US-10-197-705-356	Sequence 356, App
210	7	0.8	261	9	US-10-197-708-356	Sequence 356, App
211	7	0.8	261	9	US-10-197-699-356	Sequence 356, App
212	7	0.8	261	9	US-10-197-700-356	Sequence 356, App
213	7	0.8	261	9	US-10-197-708-356	Sequence 356, App
214	7	0.8	261	9	US-10-198-764-356	Sequence 356, App
215	7	0.8	261	9	US-10-198-765-356	Sequence 356, App
216	7	0.8	261	9	US-10-198-768-356	Sequence 356, App
217	7	0.8	261	9	US-10-198-769-356	Sequence 356, App
218	7	0.8	261	9	US-10-199-305-356	Sequence 356, App
219	7	0.8	261	9	US-10-199-306-356	Sequence 356, App
220	7	0.8	261	9	US-10-199-310-356	Sequence 356, App
221	7	0.8	261	9	US-10-199-311-356	Sequence 356, App
222	7	0.8	261	9	US-10-199-317-356	Sequence 356, App
223	7	0.8	261	9	US-10-199-314-356	Sequence 356, App
224	7	0.8	261	9	US-10-199-665-356	Sequence 356, App
225	7	0.8	261	9	US-10-199-666-356	Sequence 356, App
226	7	0.8	261	9	US-10-199-669-356	Sequence 356, App
227	7	0.8	261	9	US-10-201-534-356	Sequence 356, App
228	7	0.8	261	9	US-10-201-770-356	Sequence 356, App
229	7	0.8	261	9	US-10-201-855-356	Sequence 356, App
230	7	0.8	261	9	US-10-202-469-356	Sequence 356, App
231	7	0.8	261	9	US-10-202-470-356	Sequence 356, App
232	7	0.8	261	9	US-10-202-476-356	Sequence 356, App
233	7	0.8	261	9	US-10-202-934-356	Sequence 356, App
234	7	0.8	261	9	US-10-202-935-356	Sequence 356, App
235	7	0.8	261	9	US-10-202-936-356	Sequence 356, App
236	7	0.8	261	9	US-10-202-939-356	Sequence 356, App
237	7	0.8	261	9	US-10-205-504-356	Sequence 356, App
238	7	0.8	261	9	US-10-205-509-356	Sequence 356, App

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:35:40 ; Search time 48 Seconds
(Without alignments)
1534.152 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 919
Sequence: 1 MSKRVLEFANSPMPVALAAY.....TGVEEGNPILKHFICTPF 919

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 301932 seqs, 80129803 residues

Word size : 0

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

- Published_Applications_AA:*
- 1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
 - 2: /cgn2_6/ptodata/1/pubppaa/PC7_NEW_PUB.pep:*
 - 3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
 - 4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
 - 5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
 - 6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
 - 7: /cgn2_6/ptodata/1/pubppaa/PC7US_PUBCOMB.pep:*
 - 8: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
 - 9: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
 - 10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
 - 11: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
 - 12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
 - 13: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
 - 14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.0	851	US-10-045-792-2	Sequence 2, Appl1
2	7	0.8	41	US-10-042-141-93	Sequence 93, Appl1
3	7	0.8	41	US-09-726-643-93	Sequence 93, Appl1
4	7	0.8	51	US-10-091-504-1204	Sequence 1204, Ap
5	7	0.8	51	US-09-764-869-1204	Sequence 1204, Ap
6	7	0.8	69	US-09-809-391-601	Sequence 601, App
7	7	0.8	70	US-09-809-391-465	Sequence 465, App
8	7	0.8	95	US-09-746-919-24	Sequence 24, Appl1
9	7	0.8	102	US-09-925-297-506	Sequence 506, App
10	7	0.8	106	US-10-072-349-140	Sequence 140, App
11	7	0.8	106	US-09-764-855-140	Sequence 140, App
12	7	0.8	122	US-10-042-141-91	Sequence 91, Appl1
13	7	0.8	122	US-09-726-643-91	Sequence 1129, Ap
14	7	0.8	150	US-09-764-868-1129	Sequence 26, Appl1
15	7	0.8	158	US-09-975-132A-26	Sequence 27, Appl1
16	7	0.8	158	US-09-975-132A-27	Sequence 28, Appl1
17	7	0.8	173	US-09-975-132A-28	Sequence 368, App
18	7	0.8	187	US-09-986-480-368	Sequence 162, App
19	7	0.8	188	US-10-062-254-162	

20	7	0.8	214	US-09-764-868-956	Sequence 956, App
21	7	0.8	215	US-09-825-414-60	Sequence 60, Appl1
22	7	0.8	215	US-09-866-050A-705	Sequence 705, App
23	7	0.8	261	US-10-063-547-118	Sequence 118, App
24	7	0.8	261	US-10-174-590-356	Sequence 356, App
25	7	0.8	261	US-10-176-758-356	Sequence 356, App
26	7	0.8	261	US-10-063-616-118	Sequence 118, App
27	7	0.8	261	US-10-175-737-356	Sequence 356, App
28	7	0.8	261	US-10-063-502-118	Sequence 118, App
29	7	0.8	261	US-10-173-706-356	Sequence 356, App
30	7	0.8	261	US-10-175-738-356	Sequence 356, App
31	7	0.8	261	US-10-175-752-356	Sequence 356, App
32	7	0.8	261	US-10-176-482-356	Sequence 356, App
33	7	0.8	261	US-10-176-757-356	Sequence 356, App
34	7	0.8	261	US-10-176-913-356	Sequence 356, App
35	7	0.8	261	US-10-180-552-356	Sequence 356, App
36	7	0.8	261	US-10-180-557-356	Sequence 356, App
37	7	0.8	261	US-10-173-700-356	Sequence 356, App
38	7	0.8	261	US-10-174-572-356	Sequence 356, App
39	7	0.8	261	US-10-174-579-356	Sequence 356, App
40	7	0.8	261	US-10-174-582-356	Sequence 356, App
41	7	0.8	261	US-10-174-588-356	Sequence 356, App
42	7	0.8	261	US-10-175-739-356	Sequence 356, App
43	7	0.8	261	US-10-175-740-356	Sequence 356, App
44	7	0.8	261	US-10-176-488-356	Sequence 356, App
45	7	0.8	261	US-10-176-492-356	Sequence 356, App
46	7	0.8	261	US-10-176-747-356	Sequence 356, App
47	7	0.8	261	US-10-176-750-356	Sequence 356, App
48	7	0.8	261	US-10-176-985-356	Sequence 356, App
49	7	0.8	261	US-10-176-987-356	Sequence 356, App
50	7	0.8	261	US-10-176-991-356	Sequence 356, App
51	7	0.8	261	US-10-176-992-356	Sequence 356, App
52	7	0.8	261	US-10-176-993-356	Sequence 356, App
53	7	0.8	261	US-10-184-658-356	Sequence 356, App
54	7	0.8	261	US-10-173-695-356	Sequence 356, App
55	7	0.8	261	US-10-173-697-356	Sequence 356, App
56	7	0.8	261	US-10-173-705-356	Sequence 356, App
57	7	0.8	261	US-10-174-576-356	Sequence 356, App
58	7	0.8	261	US-10-174-585-356	Sequence 356, App
59	7	0.8	261	US-10-174-586-356	Sequence 356, App
60	7	0.8	261	US-10-175-747-356	Sequence 356, App
61	7	0.8	261	US-10-176-481-356	Sequence 356, App
62	7	0.8	261	US-10-176-485-356	Sequence 356, App
63	7	0.8	261	US-10-176-487-356	Sequence 356, App
64	7	0.8	261	US-10-176-493-356	Sequence 356, App
65	7	0.8	261	US-10-176-756-356	Sequence 356, App
66	7	0.8	261	US-10-176-911-356	Sequence 356, App
67	7	0.8	261	US-10-180-547-356	Sequence 356, App
68	7	0.8	261	US-10-176-919-356	Sequence 356, App
69	7	0.8	261	US-10-176-925-356	Sequence 356, App
70	7	0.8	261	US-10-176-978-356	Sequence 356, App
71	7	0.8	261	US-10-179-510-356	Sequence 356, App
72	7	0.8	261	US-10-180-543-356	Sequence 356, App
73	7	0.8	261	US-10-180-544-356	Sequence 356, App
74	7	0.8	261	US-10-180-546-356	Sequence 356, App
75	7	0.8	261	US-10-180-549-356	Sequence 356, App
76	7	0.8	261	US-10-180-555-356	Sequence 356, App
77	7	0.8	261	US-10-180-559-356	Sequence 356, App
78	7	0.8	261	US-10-181-000-356	Sequence 356, App
79	7	0.8	261	US-10-183-010-356	Sequence 356, App
80	7	0.8	261	US-10-183-012-356	Sequence 356, App
81	7	0.8	261	US-10-184-614-356	Sequence 356, App
82	7	0.8	261	US-10-184-623-356	Sequence 356, App
83	7	0.8	261	US-10-184-635-356	Sequence 356, App
84	7	0.8	261	US-10-184-637-356	Sequence 356, App
85	7	0.8	261	US-10-184-646-356	Sequence 356, App
86	7	0.8	261	US-10-184-647-356	Sequence 356, App
87	7	0.8	261	US-10-184-652-356	Sequence 356, App
88	7	0.8	261	US-10-187-594-356	Sequence 356, App
89	7	0.8	261	US-10-187-596-356	Sequence 356, App
90	7	0.8	261	US-10-187-745-356	Sequence 356, App
91	7	0.8	261	US-10-187-885-356	Sequence 356, App
92	7	0.8	261		

Tue Apr 29 12:22:27 2003

us-09-914-168-2.Oligo.rai

Page 24

Db 24 PLSUEEL 30

Search completed: April 28, 2003, 16:36:30
Job time : 67 secs

```

; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 64
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Mouse
US-09-724-864-64

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 264;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410
Db 76 LQAVRAL 82

RESULT 37
US-08-465-971B-2
; Sequence 2, Application US/08465971B
; Patent No. 5942414
; GENERAL INFORMATION:
; APPLICANT: YI LI and Mark D. Adams
; TITLE OF INVENTION: Human G-Protein Receptor HIBEF51
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: US
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,971B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLINS, J.G.
; REGISTRATION NUMBER: 33073
; REFERENCE/DOCKET NUMBER: 325800-453 (PFI187)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 349 amino acids
; TYPE: amino acid
; TOPOLOGY: Linear
; MOLECULE TYPE: protein
US-08-465-971B-2

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 349;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 SSOSGET 196
Db 233 SSOSGET 239

RESULT 38
US-09-134-001C-5150
; Sequence 5150, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

```

```

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GFC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5150
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5150

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 EVIDLPE 552
Db 321 EVIDLPE 327

RESULT 39
US-09-325-932A-191
; Sequence 191, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant dev
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 191
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-325-932A-191

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 355;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 752 GSSGLVS 758
Db 112 GSSGLVS 118

RESULT 40
5187089-2
; Patent No. 5187089
; APPLICANT: SCOTT, RANDY W.; GOLINI, FRED; MCGROGAN, MICHAEL
; TITLE OF INVENTION: PROTEASE NEXIN-I VARIANTS WHICH INHIBIT
; ELASTASE
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/542,484
; FILING DATE: 21-JUN-1990
; SEQ ID NO: 2
; LENGTH: 397
; 5187089-2

Query Match
Best Local Similarity 100.0%; Score 7; DB 6; Length 397;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 PLSTBEL 123

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/245,511
; FILING DATE: 18-MAY-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/116,541
; FILING DATE: 01-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-069 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 175 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; ORGANISM: Neisseria gonorrhoeae
; US-08-245-511-26

```

```

Query Match
Best Local Similarity 0.8%; Score 7; DB 2; Length 175;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 354 DTGTOYR 360
Db 11 DTGTOYR 17

```

```

RESULT 34
US-08-600-993A-26
; Sequence 26, Application US/08600993A
; Patent No. 5981229
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomanen, Elaine
; TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
; TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600,993A
; FILING DATE: 1-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/245,511
; FILING DATE: 18-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/116,541
; FILING DATE: 01-SEP-1994

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-069 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 175 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; ORGANISM: Neisseria gonorrhoeae
; US-08-600-993A-26

```

```

Query Match
Best Local Similarity 0.8%; Score 7; DB 2; Length 175;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 354 DTGTOYR 360
Db 11 DTGTOYR 17

```

```

RESULT 35
US-09-323-872A-12
; Sequence 12, Application US/09323872A
; Patent No. 6395539
; GENERAL INFORMATION:
; APPLICANT: Coschigano, Peter
; TITLE OF INVENTION: Compositions and Methods for Bioremediation
; FILE REFERENCE: OHU-03640
; CURRENT APPLICATION NUMBER: US/09/323,872A
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/072,433
; PRIOR FILING DATE: 1998-05-04
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 12
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Azorhizobium caulinodans
; US-09-323-872A-12

```

```

Query Match
Best Local Similarity 0.8%; Score 7; DB 4; Length 211;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 721 ALAGVA 727
Db 197 ALAGVA 203

```

```

RESULT 36
US-09-724-864-64
; Sequence 64, Application US/09724864
; Patent No. 6380362
; GENERAL INFORMATION:
; APPLICANT: Watson, James D
; APPLICANT: Murison, James G.
; TITLE OF INVENTION: Polynucleotides, polypeptides expressed
; TITLE OF INVENTION: by the polynucleotides and methods for their use.
; FILE REFERENCE: 11000,105001
; CURRENT APPLICATION NUMBER: US/09/724,864
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678
; PRIOR FILING DATE: 1999-12-23

```

Db 42 LLEOLLT 48

RESULT 31
US-09-045-467-24
Sequence 24, Application US/09045467
GENERAL INFORMATION:
APPLICANT: Johnson, Howard M.
Ponzar, Carol H.
TITLE OF INVENTION: Interferon Tau Compositions and
Methods of Use
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Ave., Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/045,467
FILING DATE: 20-Mar-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/455,021
FILING DATE: 31-MAY-1995
APPLICATION NUMBER: US 08/438,753
FILING DATE: 10-MAY-1995
APPLICATION NUMBER: US 08/139,891
FILING DATE: 19-OCT-1993
APPLICATION NUMBER: US 07/847,741
FILING DATE: 09-MAR-1992
APPLICATION NUMBER: US 07/318,050
FILING DATE: 02-MAR-1989
APPLICATION NUMBER: US 07/969,890
FILING DATE: 30-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Dehlinger, Peter J.
REGISTRATION NUMBER: 28,006
REFERENCE/DOCKET NUMBER: 5600-0001.36
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-0880
TELEFAX: 650-324-0960
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: predicted amino acid coding sequence
of SEQ ID NO:23 (Hulfntau7).
SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-09-045-467-24

Query Match 0.8%; Score 7; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 389 LLEOLLT 395
|||||
Db 42 LLEOLLT 48

RESULT 32
US-08-796-792-2
Sequence 2, Application US/08796792
Patent No. 6087163

GENERAL INFORMATION:
APPLICANT: Gennaro, Maria L.
APPLICANT: Lyashchenko, Konstantin P.
APPLICANT: Manca, Claudia M.A.
TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS
TITLE OF INVENTION: SPECIFIC PROTEINS AND GENES, MIXTURES OF ANTIGENS AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 45 Rockefeller Plaza, Suite 2800
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/796,792
FILING DATE: 06-FEB-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/011,364
FILING DATE: 09-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Hone, William J.
REGISTRATION NUMBER: 26,739
REFERENCE/DOCKET NUMBER: 07763/03301
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-765-5070
TELEFAX: 212-258-2291
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-796-792-2

Query Match 0.8%; Score 7; DB 3; Length 159;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 PVALAAV 20
|||||
Db 25 PVALAAV 31

RESULT 33
US-08-245-511-26
Sequence 26, Application US/08245511
Patent No. 5928900
GENERAL INFORMATION:
APPLICANT: Masure, H Robert
APPLICANT: Pearce, Barbara J
APPLICANT: Tuomenen, Elaine
TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

Query Match 0.8%; Score 7; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 389 LLEQLLT 395
Db 42 LLEQLLT 48

Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 389 LLEQLLT 395
Db 42 LLEQLLT 48

RESULT 29
US-08-455-524B-24
; Sequence 24, Application US/08455524B
; Patent No. 5942223
; GENERAL INFORMATION:
; APPLICANT: Bazer, Fuller W.
; APPLICANT: Johnson, Howard M.
; APPLICANT: Pontzer, Carol H.
; APPLICANT: Olt, Troy L.
; APPLICANT: Van Heeke, Gino
; TITLE OF INVENTION: Interferon Tau Compositions and
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455,524B
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/438,753
; FILING DATE: 10-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/139,891
; FILING DATE: 19-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/847,741
; FILING DATE: 09-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/318,050
; FILING DATE: 02-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/969,890
; FILING DATE: 30-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 5600-0001.32
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-324-0880
; TELEFAX: 415-324-0960
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: predicted amino acid coding sequence
; INDIVIDUAL ISOLATE: of SEQ ID NO:23 (HUIFNTAU7).
US-08-455-524B-24
Query Match 0.8%; Score 7; DB 2; Length 95;

RESULT 30
US-08-455-021B-24
; Sequence 24, Application US/08455021B
; GENERAL INFORMATION:
; APPLICANT: Bazer, Fuller W.
; APPLICANT: Johnson, Howard M.
; APPLICANT: Pontzer, Carol H.
; APPLICANT: Olt, Troy L.
; APPLICANT: Van Heeke, Gino
; APPLICANT: Imakawa, Kazuhito
; TITLE OF INVENTION: Interferon Tau Compositions and
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455,021B
; FILING DATE: 31-MAY-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/139,891
; FILING DATE: 19-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/847,741
; FILING DATE: 09-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/318,050
; FILING DATE: 02-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/969,890
; FILING DATE: 30-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 5600-0001.31
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-324-0880
; TELEFAX: 415-324-0960
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: predicted amino acid coding sequence
; INDIVIDUAL ISOLATE: of SEQ ID NO:23 (HUIFNTAU7).
US-08-455-021B-24
Query Match 0.8%; Score 7; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 389 LLEQLLT 395
Db 42 LLEQLLT 48

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/969,890
FILING DATE: 30-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 5600-0001.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-324-0880
TELEFAX: 415-324-0960
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: predicted amino acid coding sequence
US-08-438-753B-24
INDIVIDUAL ISOLATE: of SEQ ID NO:23 (HuIFNtau7).

Query Match 0.8%; Score 7; DB 1; Length 95;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 389 LLEOLLT 395
|||||
DB 42 LLEOLLT 48

RESULT 27

US-08-443-883A-24

Sequence 24, Application US/08443883A

Patent No. 5738845

GENERAL INFORMATION:

APPLICANT: Bazer, Fuller W.

APPLICANT: Johnson, Howard M.

APPLICANT: Pontzer, Carol H.

APPLICANT: Ott, Troy L.

APPLICANT: Van Hecke, Gino

APPLICANT: Imakawa, Kazuhito

TITLE OF INVENTION: Interferon Tau Compositions and

TITLE OF INVENTION: Methods of Use

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:

ADDRESS: Dehlinger & Associates

STREET: 350 Cambridge Ave., Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/443,883A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/139,891

FILING DATE: 19-OCT-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/847,741

FILING DATE: 09-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/318,050

FILING DATE: 02-MAR-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/969,890

FILING DATE: 30-OCT-1992

ATTORNEY/AGENT INFORMATION:

NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 5600-0001.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-324-0880
TELEFAX: 415-324-0960
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: predicted amino acid coding sequence
US-08-443-883A-24
INDIVIDUAL ISOLATE: of SEQ ID NO:23 (HuIFNtau7).

Query Match 0.8%; Score 7; DB 1; Length 95;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 389 LLEOLLT 395
|||||
DB 42 LLEOLLT 48

RESULT 28

US-08-631-328-24

Sequence 24, Application US/08631328

Patent No. 5939266

GENERAL INFORMATION:

APPLICANT: Johnson, Howard M.

APPLICANT: Pontzer, Carol H.

APPLICANT: Subramaniam, Prem S.

TITLE OF INVENTION: Hybrid Interferon Compositions and

TITLE OF INVENTION: Methods of Use

NUMBER OF SEQUENCES: 55

CORRESPONDENCE ADDRESS:

ADDRESS: Dehlinger & Associates

STREET: 350 Cambridge Ave., Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/631,328

FILING DATE: 12-APR-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/438,753

FILING DATE: 10-MAY-1995

ATTORNEY/AGENT INFORMATION:

NAME: Sholtz, Charles K.

REGISTRATION NUMBER: 38,615

REFERENCE/DOCKET NUMBER: 5600-0001.34

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-324-0880

TELEFAX: 415-324-0960

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:

LENGTH: 95 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: predicted amino acid coding sequence

US-08-631-328-24

OC	NCBI_TaxID=562, 83334; X	OS	Escherichia coli, O157:H7.	OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
DE	YTFM OR B4220 OR 25831 OR EC55198.	NCBI_TaxID=562, 83334; X <td>Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td> </td>	Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td>	OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
GN	YTFM OR B4220 OR 25831 OR EC55198.	NCBI_TaxID=562, 83334; X <td>Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td> </td>	Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td>	OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
DT	16-OCT-2001 (Rel. 40, Last annotation update)	NCBI_TaxID=562, 83334; X <td>Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td> </td>	Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td>	OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
DT	01-FEB-1995 (Rel. 31, Last sequence update)	NCBI_TaxID=562, 83334; X <td>Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td> </td>	Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td>	OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
AC	P39320;	NCBI_TaxID=562, 83334; X <td>Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td> </td>	Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td>	OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
ID	YTFM_ECOLI	NCBI_TaxID=562, 83334; X <td>Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td> </td>	Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td>	OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
RESULT 5	YTFM_ECOLI	NCBI_TaxID=562, 83334; X <td>Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td> </td>	Escherichia coli, O157:H7. <td>OC</td> <td>Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.</td>	OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

RESULT	6
MCCA_ARATH	
ID	STANDARD:
MCCA_ARATH	PRT: 734 AA.
AC	Q42523; Q9SA61;
DT	15-JUN-2002 (Rel. 41, Created)
DT	15-JUN-2002 (Rel. 41, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Methylcrotonyl-CoA carboxylase alpha chain, mitochondrial precursor (EC 6.4.1.4) (3-methylcrotonyl-CoA carboxylase 1) (MCase alpha subunit)
DE	subunit1 (3-methylcrotonyl-CoA:carbon dioxide lyase alpha subunit)

GN MCCA OR AT1G03090 OR F1003_8 OR F1003.9.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 ON NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=95232183; PubMed=7716229;
 RA Weaver L.M., Lebrun L., Franklin A., Huang L., Hoffman N.,
 RT Wurttele E.S., Nikolau B.J.;
 RL "Molecular cloning of the biotinylated subunit of 3-methylcrotonyl-
 coenzyme A carboxylase of Arabidopsis thaliana.";
 RL Plant Physiol. 107:1013-1014(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
 RA Dunn P., Elgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malti R., Marzilli A.,
 RA Millschler J., Miranda M., Nguyen M., Nierman W.C., Osborne B.T.,
 RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salsberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Uteback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RL "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:816-820(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/PBEC).";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP TISSUE SPECIFICITY.
 RC STRAIN=cv. Landsberg erecta, and cv. Columbia;
 RX MEDLINE=20148760; PubMed=1061539;
 RA Mckean A.L., Ke J., Song J., Che P., Achenbach S., Nikolau B.J.,
 RA Wurttele E.S.;
 RL "Molecular characterization of the non-biotin-containing subunit of
 3-methylcrotonyl-CoA carboxylase.";
 RL J. Biol. Chem. 275:5582-5590(2000).
 CC -1- CATALYTIC ACTIVITY: ATP + 3-methylcrotonyl-CoA + HCO(3)(-) = ADP +
 phosphate + 3-methylglutacetyl-CoA.
 CC -1- COFACTOR: Biotin.
 CC -1- PATHWAY: Leucine catabolism.
 CC -1- SUBUNIT: Probably a heterodimer composed of biotin-containing
 alpha subunits and beta subunits (By similarity).
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -1- TISSUE SPECIFICITY: In roots, cotyledons, leaves, flowers,
 ovaries, siliques and embryos.
 CC -1- MISCELLANEOUS: Temporal and spatial accumulation of the alpha and
 beta subunits during development at approximately equal molar
 ratios.
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous
 gene model prediction.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U12536; AAA67356.1; -;
 DR EMBL; AC006550; AAD25800.1; ALT_SEQ.
 DR EMBL; AY070723; AAL50065.1; -;
 DR HSSP; P24182; 1BNC.
 DR InterPro; IPR001882; Biotin_attach.
 DR InterPro; IPR000089; Biotin_lipoyl.
 DR InterPro; IPR000901; CPSase.
 DR Pfam; PF02785; Biotin_card_C; 1.
 DR Pfam; PF00364; Biotin_lipoyl; 1.
 DR Pfam; PF00289; CPSase_L_chain; 2.
 DR Pfam; PF02786; CPSase_L_D2; 2.
 DR PROSITE; PS00188; BIOTIN; 1.
 DR PROSITE; PS00866; CPSASE_1; 1.
 DR PROSITE; PS00867; CPSASE_2; 1.
 DR PROSITE; PS00867; CPSASE_2; 1.
 KW Mitochondrion; Ligase; Biotin; ATP-binding; Transit peptide.
 FT TRANSIT 1 25
 FT CHAIN 26 734
 FT NP_BIND 198 203
 FT ACT_SITE 329 329
 FT BINDING 699 699
 FT CONFLICT 85 85
 FT CONFLICT 92 92
 FT CONFLICT 281 300
 FT CONFLICT 430 430
 FT SEQUENCE 734 AA; 80451 MW; 251CACF646B046B CRC64;
 SO
 Query Match 0.9%; Score 8; DB 1; Length 734;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 241 AARAVGY 248
 DB 298 AARAVGY 305
 |||||
 DB 298 AARAVGY 305
 RESULT 7
 CARB_THEAC STANDARD; PTR; 1047 AA.
 ID CARB_THEAC
 AC Q9HK17;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
 DE phosphate synthetase ammonia chain).
 GN CARB OR TA0791.
 OS Thermoplasma acidophilum.
 CC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 CC Thermoplasmataceae; Thermoplasma.
 CC NCBI_TaxID=2303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
 RT acidophilum.";
 RL Nature 407:508-513(2000).
 CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
 phosphate + L-glutamate + carbamoyl phosphate.
 CC -1- COFACTOR: Binds three manganese ions (By similarity).
 CC -1- PATHWAY: Arginine biosynthesis.
 CC -1- PATHWAY: Pyrimidine biosynthesis: first step.
 CC -1- SUBUNIT: Composed of two chains: the small (or glutamine) chain
 CC promotes the hydrolysis of glutamine to ammonia, which is used by
 CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
 CC similarity).
 CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL445065; CAC11922.1; -
DR HSSP: P00968; LJDB.
DR InterPro: IPR005483; CPase_L.
DR InterPro: IPR005479; CPase_L_D2.
DR InterPro: IPR005480; CPase_L_D3.
DR InterPro: IPR005481; CPase_L_N.
DR InterPro: IPR004362; MGS_Like.
DR Pfam: PF00289; CPase_L_D2; 1.
DR Pfam: PF02786; CPase_L_D3; 1.
DR Pfam: PF02787; CPase_L_D3; 1.
DR Pfam: PF02142; MGS; 1.
DR PRINTS: PR00098; CPASE.
DR PROSITE: PS00866; CPASE_1; FALSE_NEG.
DR PROSITE: PS00867; CPASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 398 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 399 539 OLIGOMERIZATION DOMAIN.
FT DOMAIN 540 915 CARAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 916 1047 ALLOSTERIC DOMAIN.
FT REPEAT 1 539
FT REPEAT 540 1047
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 300 350
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 296 296 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 2 (BY SIMILARITY).
FT METAL 813 813 MANGANESE 3 (BY SIMILARITY).
FT METAL 825 825 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1047 AA; 116851 MW; E062CAD131746864 CRC64;

Query Match 0.9%; Score 8; DB 1; Length 1047;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 626 KGVKLYAT 633
DB 941 KGVKLYAT 948

RESULT 8
SALM_DROME STANDARD: PRT: 1355 AA.
ID SALM_DROME
AC P39770;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeotic protein spalt-major.
DE SALM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94139659; PubMed=7905822;
RA Kuehlein R.P., Frommer G., Friedrich M., Gonzalez-Gaitan M.,
RA Weber A., Wagner-Bernholz J.F., Gehring W.J., Jaekle H., Schuh R.;
RT "spalt encodes an evolutionarily conserved zinc finger protein of
RT novel structure which provides homeotic gene function in the head and
RT tail region of the Drosophila embryo.";
RL EMBL J. 13:168-179(1994).
CC -!- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF THE POSTERIOR-MOST

```

```

CC HEAD AND THE ANTERIOR-MOST TAIL SEGMENTS OF THE EMBRYO. PROBABLY
CC FUNCTION AS A TRANSCRIPTIONAL REGULATOR. COULD REPRESS THE
CC TRANSCRIPTION OF THE TSH GENE.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable)
CC -!- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT BLASTODERM STAGE AND
CC LATER IN RESTRICTED AREAS OF THE EMBRYONIC NERVOUS SYSTEM AS
CC WELL AS IN THE DEVELOPING TRACHEA.
CC -!- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X7541; CAA53229.1; -
DR HSSP: P15822; IBBO.
DR Flybase: FBgn004579; salm.
DR InterPro: IPR000822; Znf.C2H2.
DR Pfam: PF00096; zfc-C2H2; 7.
DR PRINTS: PR00048; ZINCFINGER.
DR SMART: SM00355; Znf.C2H2; 7.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.
KW Developmental protein; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Transcription regulation; Repeat.
FT ZN_FING 451 473 C2H2-TYPE.
FT ZN_FING 479 501 C2H2-TYPE.
FT ZN_FING 824 846 C2H2-TYPE.
FT ZN_FING 852 874 C2H2-TYPE.
FT ZN_FING 884 906 C2H2-TYPE.
FT ZN_FING 1289 1311 C2H2-TYPE.
FT ZN_FING 1317 1339 C2H2-TYPE.
SQ SEQUENCE 1355 AA; 148995 MW; 6FC4EBDC0BC6355E CRC64;

Query Match 0.9%; Score 8; DB 1; Length 1355;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LRPEDIOA 68
DB 913 LRPEDIOA 920

RESULT 9
SALM_DROVI STANDARD: PRT: 1402 AA.
ID SALM_DROVI
AC P39606;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeotic protein spalt-major.
DE SALM.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7244;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94139659; PubMed=7905822;
RA Kuehlein R.P., Frommer G., Friedrich M., Gonzalez-Gaitan M.,
RA Weber A., Wagner-Bernholz J.F., Gehring W.J., Jaekle H., Schuh R.;
RT "spalt encodes an evolutionarily conserved zinc finger protein of
RT novel structure which provides homeotic gene function in the head and
RT tail region of the Drosophila embryo.";
RL EMBL J. 13:168-179(1994).
CC -!- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF THE POSTERIOR-MOST
CC HEAD AND THE ANTERIOR-MOST TAIL SEGMENTS OF THE EMBRYO. PROBABLY
CC FUNCTION AS A TRANSCRIPTIONAL REGULATOR. COULD REPRESS THE

```

```

CC TRANSCRIPTION OF THE TSH GENE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -2- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT BLASTODERM STAGE AND
CC LATER IN RESTRICTED AREAS OF THE EMBRYONIC NERVOUS SYSTEM AS
CC WELL AS IN THE DEVELOPING TRACHEA.
CC -3- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z27444; CAA81800.1; -
DR PIR: S42748; S42748.
DR HSSP: P15822; 1BBO.
DR FLYBASE: FBgn0013137; Dvir\salin.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2.7.
DR PRINTS: PR00048; ZINC_FINGER.
DR SMART: SM00355; Znf_C2H2.7.
DR PROSITE: PS00026; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.
KW Developmental protein; zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Transcription regulation; Repeat.
FT ZN_FING 432 454 C2H2-TYPE.
FT ZN_FING 460 482 C2H2-TYPE.
FT ZN_FING 845 867 C2H2-TYPE.
FT ZN_FING 873 895 C2H2-TYPE.
FT ZN_FING 905 927 C2H2-TYPE.
FT ZN_FING 1336 1358 C2H2-TYPE.
FT ZN_FING 1364 1386 C2H2-TYPE.
SO SEQUENCE 1402 AA; 153745 MW; 4CDB2E384F70A5 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 1402;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 LTPRO10A 68
DB 934 LTPRO10A 941

RESULT 10
HYBG_ECOLI STANDARD; PRT; 82 AA.
ID HYBG_ECOLI
AC P37185;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hydrogenase-2 operon protein hydg.
GN HYBG OR B2990 OR Z4344 OR ECS3875.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / TG1;
RX MEDLINE=94292472; PubMed=8021226;
RA Menon N.K., Chatelus C.Y., Deravartanian M., Wendt J.C.,
RA Shammugam K.T., Peck H.D.Jr., Przybyla A.E.;
RT "Cloning, sequencing, and mutational analysis of the hyd operon
RT encoding Escherichia coli hydrogenase 2.";
RL J. Bacteriol. 176:4416-4423(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97426617; PubMed=9278503;

```

```

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocitis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhnara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1- FUNCTION: MAY HAVE A SPECIFIC ROLE IN THE MATURATION OF THE LARGE
CC SUBUNIT OF HYD1 AND HYD2.
CC -2- SIMILARITY: BELONGS TO THE HUPE/HYPC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U09177; AAA21595.1; -
DR EMBL: U28377; AAA69157.1; -
DR EMBL: AE000382; AAC76026.1; -
DR EMBL: AE005529; AAG58127.1; -
DR EMBL: AP002563; BAB37298.1; -
DR Ecogen: EG11805; hydg.
DR InterPro: IPR001109; HUPE_HYPC.
DR Pfam: PF01455; HUPE_HYPC.1.
DR PRINTS: PR00445; HUPEHYPC.
DR PRODOM: PD003112; HUPE_HYPC.1.
DR TIGRPFMS: TIGR00074; hupC_hupE.1.
DR PROSITE: PS01097; HUPE_HYPC.1.
KW Complete proteome.
SQ SEQUENCE 82 AA; 8808 MW; 8E9736BFC4EFD626 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 839 GOVLAVG 845
DB 7 GOVLAVG 13

RESULT 11
YGT2_YEAST STANDARD; PRT; 107 AA.
ID YGT2_YEAST
AC P53103;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 12.2 kDa protein in COX4-GT51 intergenic region.

```

```

GN YG182C OR G1607.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:S288C / FY1679;
RX MEDLINE=97197971; PubMed=9046087;
RA Coglievina M., Klima R., Bertani I., Delneri D., Zaccaria P.,
RA Brusch C.V.;
RT "Sequencing of a 40.5 kb fragment located on the left arm of
RT chromosome VII from Saccharomyces cerevisiae.";
RL Yeast 13:55-64(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X91489; CAA62792.1;
DR EMBL: Z72704; CAA96894.1;
DR SGD: S0003150; YG182C.
KW Hypothetical protein.
SO SEQUENCE 107 AA; 12220 MW; 1D22E64BC819A499 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 687 YSLRYRL 693
   |||||
DB 84 YSLRYRL 90

RESULT 12
Y353_MYCPN STANDARD; PRT; 109 AA.
AC P75249;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG353 homolog (G12-ori109).
GN MPM529 OR MP13.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirk E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000028; AAB95961.1;
DR InterPro: IPR000119; Bac_DNABind.
DR SMART: SM00411; BHL. 1
KW Hypothetical protein; Complete proteome.
SO SEQUENCE 109 AA; 12410 MW; 3FF862745F3C5316 CRC64;

```

```

Query Match 0.8%; Score 7; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 898 VATGYKE 904
   |||||
DB 22 VATGYKE 28

RESULT 13
MP63_MYCTU STANDARD; PRT; 159 AA.
AC P97175; 008224;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Immunogenic protein MP63/MPB63 precursor (Antigen MP63/MPB63) (16
DE kDa immunoprotective extracellular protein).
GN MP63 OR MPB63 OR RV1926C OR MT1977 OR MTCY09F9.38.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-M.tuberculosis; STRAIN-H37Rv;
RX MEDLINE=97130011; PubMed=8975887;
RA Mance C.M.A., Lyashchenko K., Wiker H.G., Usai D., Colangeli R.,
RA Genaro M.L.;
RT "Molecular cloning, purification, and serological characterization of
RT MP63, a novel antigen secreted by Mycobacterium tuberculosis.";
RL Infect. Immun. 65:16-23(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-M.tuberculosis; STRAIN-Erdmann;
RX MEDLINE=97313166; PubMed=9169770;
RA Harth G., Lee B.Y., Horwitz M.A.;
RT "High-level heterologous expression and secretion in rapidly growing
RT nonpathogenic mycobacteria of four major Mycobacterium tuberculosis
RT extracellular proteins considered to be leading vaccine candidates
RT and drug targets";
RL Infect. Immun. 65:2321-2328(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-M.tuberculosis; STRAIN-H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Horsby T., Jagels K., Krogh A., Molean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES-M.tuberculosis; STRAIN-CDC 1551 / Oshkosh;
RX Fletschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Debby R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA DeLoney A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE OF 30-49.
RC SPECIES-M.tuberculosis;

```



```

RX MEDLINE=95183499; Pubmed-7878014;
RA Horwitz M.A., Lee B.W., Dillon B.J., Harth G.;
RT "Protective immunity against tuberculosis induced by vaccination with
RT major extracellular proteins of Mycobacterium tuberculosis.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:1530-1534(1995).
RN [6]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC SPECIES=M.tuberculosis;
RA MEDLINE=91099899; Pubmed-1898899;
RA Nagai S., Wiker H.G., Harboe M., Kinomoto M.;
RT "Isolation and partial characterization of major protein antigens in
RT the culture fluid of Mycobacterium tuberculosis.";
RL Infect. Immun. 59:372-382(1991).
RN [7]
RP SEQUENCE FROM N.A.
RC SPECIES=M.dovisi; STRAIN=BCG / Tokyo;
RA Kamile K., Matsuda S., Kobayashi A., Kobayashi K.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U27119; AAB95083.1; -
CC EMBL: U82234; AAB61537.1; -
CC EMBL: Z84498; CAB06500.1; -
CC EMBL: AE007052; AAK46249.1; -
CC EMBL: AB048799; BAB39210.1; -
CC TIGR: MT1977; -
CC Tuberculin: RV1926c; -
CC Antigen: Signal; Complete proteome.
CC SIGNAL 1 29
CC CHAIN 30 159
CC SEQUENCE 159 AA; 16514 MW; EB563E1EABCE428 CRC64;
SQ
Query Match 0.8%; Score 7; DB 1; Length 159;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 PVALAAY 20
DB 25 PVALAAY 31
RESULT 14
ID GRAL_TOXGO STANDARD; PRT; 190 AA.
AC P13403;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dense granule protein 1 precursor (Protein GRA 1) (Major antigen
DE P24).
DE GN GRAL.
OS Toxoplasma gondii.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;
OC Toxoplasma.
OX NCBI_Taxid=5811;
OX [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90017513; Pubmed=2798425;
RA Cesbron-Delauw M.F., Guy B., Torprier G., Pierce R.J., Lenzen G.,
RA Cesbron J.Y., Charif H., Lepage P., Darcy F., Lecocq J.P.,
RA Capron A.;
RT "Molecular characterization of a 23-kilodalton major antigen secreted
RT by Toxoplasma gondii.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7537-7541(1989).
CC -1- SUBCELLULAR LOCATION: LOCATED IN DENSE GRANULES OF TACHYZOITES.

```

```

CC -1- CAUTION: ACCORDING TO REF.1 THERE ARE TWO POTENTIAL EF-HAND
CC REGIONS. BUT THE EVIDENCE SEEMS WEAK.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M26007; AAA30141.1; -
CC PIR: A33839; A33839.
CC Antigen: Signal.
CC SIGNAL 1 24
CC CHAIN 25 190
CC CARBHYD 30 30
CC SEQUENCE 190 AA; 20149 MW; 3B05D2610C615A53 CRC64;
SQ
Query Match 0.8%; Score 7; DB 1; Length 190;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 529 DAVSAVA 535
DB 149 DAVSAVA 155
RESULT 15
ID RS7_AVIMR STANDARD; PRT; 190 AA.
AC Q92NS1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 40S ribosomal protein S7.
DE RS7.
GN Avicennia marina.
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Eukaryota; Viridiplantae; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Avicenniaceae; Avicennia.
OX NCBI_Taxid=82927;
OX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RA Parani M., Senthilkumar P., Lakshmi M., Parida A.;
RT "A nuclear gene from Avicennia marina homologous to the 40S ribosomal
RT protein S7 gene of animals and human.";
RL (in) Plant Gene Register PGR8-203.
CC -1- SIMILARITY: BELONGS TO THE S7E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF056316; AAD03501.1; -
CC EMBL: AF098519; AAC97947.1; -
CC InterPro: IPR000554; Ribosomal_S7E.
CC Pfam: PF01251; Ribosomal_S7e; 1.
CC ProDom: PD006276; Ribosomal_S7e; 1.
CC PROSITE: PS00948; RIBOSOMAL_S7E; FALSE-NEG.
CC RIBOSOMAL protein.
CC SEQUENCE 190 AA; 21871 MW; CA5173ACF57A47B0 CRC64;
SQ
Query Match 0.8%; Score 7; DB 1; Length 190;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 804 HVPYRLR 810

```

DB 64 HYPHRLR 70

|||||

RESULT 16
ID INO1_BOVIN STANDARD: PRT: 195 AA.

AC P07352:
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Interferon omega-1 precursor (Interferon alpha-II-1) (IFN-omega-c1).
GN IFNM.

OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE-Pancreas;
RX MEDLINE=85187974; PubMed=2985969;
RA Capon D.J., Shepard H.M., Goeddel D.V.;
RT "Two distinct families of human and bovine interferon-alpha genes are
RT coordinately expressed and encode functional polypeptides."
RL Mol. Cell. Biol. 5:768-779(1985).
RN [2]

RP SEQUENCE FROM N.A.
RA Chung Y.G., Seidel G.E. Jr.;
RT "Cloning bovine interferon-tau genes and characterizing their
RT transcriptional expression during early pregnancy."
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INTERFERON ALPHA, BETA AND DELTA
CC FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: M11002; AAA30578.1; -;
DR EMBL: AF238610; AAG14167.1; -;
DR PIR: B23385; IYPOIT.

DR HSSP: P01563; ZHIE.

DR InterPro: IPR000471; Interferon_abd.

DR Pfam: PF00143; Interferon_1.

DR PRINTS: PR00266; INTERFERONAB.

DR ProDom: PD000550; Interferon_abd. 1.

DR SMART: SM00076; IFabd. 1.

DR PROSITE: PS00252; INTERFERON_A-B_D; 1.

KW Cytokine; Antiviral; Multigene family; Signal.

FT SIGNAL 1 23

FT CHAIN 24 195 INTERFERON OMEGA-1.

FT DISULFID 24 122 BY SIMILARITY.

FT DISULFID 52 162 BY SIMILARITY.

SO SEQUENCE 195 AA; 21980 MW; 1B65FD4BCABB2A15 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 195;

Best Local Similarity 100.0%; Pred. No. 42;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 389 LEQQLT 395

DB 104 LLEQLLT 110

RESULT 17

FD01_ECOLI STANDARD: PRT: 211 AA.

ID FD01_ECOLI

P32174;

AC 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Formate dehydrogenase, cytochrome b556(FD0) subunit (Formate

dehydrogenase-O gamma subunit) (FDH-Z gamma subunit) (Aerobic formate

dehydrogenase cytochrome b556 subunit).

GN FD01 OR B3892 OR Z5434 OR ECS4818.

OS Escherichia coli, and

OC Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562, 83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE=93347969; PubMed=8346018;

RA Plunkett G., III, Burland V.D., Daniels D.L., Blattner F.R.;

RT "Analysis of the Escherichia coli genome. III. DNA sequence of the

RT region from 87.2 to 89.2 minutes."

RL Nucleic Acids Res. 21:3391-3398(1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."

RT Nature 409:529-533(2001).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-O157:H7 / RIMD 0509952;

RX MEDLINE=21156331; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,

RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,

RA Kikura S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohaemorrhagic Escherichia coli

RT O157:H7 and genomic comparison with a laboratory strain K-12."

RL DNA Res. 8:11-22(2001).

RN [4]

RP CHARACTERIZATION.

RC STRAIN-K12;

RX MEDLINE=96099298; PubMed=8522521;

RA Abalou H., Pommer J., Giordano G., Mandrand-Berthelet M.-A.;

RT "Expression and characterization of the Escherichia coli fd0 locus

RT and a possible physiological role for aerobic formate

RT dehydrogenase."

RL J. Bacteriol. 177:7141-7149(1995).

RN [5]

RP TOPOLOGY.

RX MEDLINE=99069338; PubMed=9852007;

RA Benoit S., Abalou H., Mandrand-Berthelet M.-A.;

RT "Topological analysis of the aerobic membrane-bound formate

RT dehydrogenase of Escherichia coli."

RL J. Bacteriol. 180:6625-6634(1998).

CC -1- FUNCTION: ALLOWS TO USE FORMATE AS MAJOR ELECTRON DONOR DURING
CC AEROBIC RESPIRATION. SUBUNIT GAMMA IS PROBABLY THE CYTOCHROME
CC B556(FD0) COMPONENT OF THE FORMATE DEHYDROGENASE.
CC -1- SUBUNIT: FORMATE DEHYDROGENASE IS A MEMBRANE-BOUND COMPLEX, FORMED
CC BY SUBUNITS ALPHA, BETA AND GAMMA.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: STRONG, TO PDN1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; L19201; AAB03025.1; -
DR EMBL; AE000464; AAD13454.1; -
DR EMBL; AE005620; AAG59085.1; -
DR EMBL; AP002567; BAB38241.1; -
DR PIR; S40836; S40836.
KW Ecodene; E61856; fdoi.
DM Electron transport; Heme; Iron; Transmembrane; Complete proteome.
FT DOMAIN 1 17 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 18 32 PROBABLE.
FT DOMAIN 33 53 PERIPLASMIC (PROBABLE).
FT TRANSMEM 54 72 PROBABLE.
FT DOMAIN 73 112 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 113 130 PERIPLASMIC (PROBABLE).
FT DOMAIN 131 151 PERIPLASMIC (PROBABLE).
FT TRANSMEM 152 170 PROBABLE.
FT DOMAIN 171 211 CYTOPLASMIC (PROBABLE).
SQ SEQUENCE 211 AA; 24606 MW; 166AC5A661C738D5 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 HNLNRD 611
DB 78 HNLNRD 84
IIIIII

RESULT 18
FIXJ_AZCCA STANDARD; PRT; 211 AA.
AC P26487;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulatory protein fixJ.
GN FIXJ.
OS Azorhizobium caulinodans.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Hyphomicrobium group; Azorhizobium.
OX NCBI_TaxID=7;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORS571;
RX MEDLINE=91260451; PubMed=2046550;
RA Kaminski P.A., Elmerich C.;
RT *Involvement of fixJ in the regulation of nitrogen fixation in
RT Azorhizobium caulinodans.";
RL Mol. Microbiol. 5:665-673(1991).
CC -!- FUNCTION: FIXJ, WHEN ACTIVATED BY FIXL, INDUCES THE EXPRESSION OF
CC BOTH NIF, REQUIRED FOR ACTIVATION OF CLASSICAL NIF AND FIX GENES,
CC AND FIXK, REQUIRED FOR FIXN ACTIVATION.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- PTM: PHOSPHORYLATED BY FIXL (PROBABLE).
CC -!- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; X56658; CAA39980.1; -
DR PIR; S15167; S15167.
DR HSSP; P10958; IDBM.
DR InterPro; IPR000792; HTH_LuxR.
DR InterPro; IPR001789; Response_reg.

DR Pfam; PF00072; response_reg; 1.
DR Pfam; PF00196; GerE; 1.
DR PRINTS; PR00038; HTHLUXR.
DR ProDom; PD000039; Response_reg; 1.
DR ProDom; PD000307; HTH_LuxR; 1.
DR SMART; SM00421; HTH_LUXR; 1.
DR SMART; SM00448; REC; 1.
DR PROSITE; PS00622; HTH_LUXR_FAMILY; 1.
DR PROSITE; PS50110; RESPONSE REGULATORY; 1.
KW Sensory transduction; Phosphorylation; Transcription regulation;
KW DNA-binding; Activator; Nitrogen fixation
FT DOMAIN 6 120 RESPONSE REGULATORY.
FT MOD_RES 55 55 PHOSPHORYLATION (BY SIMILARITY).
FT DNA_BIND 160 179 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 211 AA; 22282 MW; E0F06B3BA4CD827D CRC64;

Query Match 0.8%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 ALLAGVA 727
DB 197 ALLAGVA 203
IIIIII

RESULT 19
VATD_THERH STANDARD; PRT; 223 AA.
ID VATD_THERH
AC 087880;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE V-type ATP synthase subunit D (BC 3.6.3.14) (V-type ATPase subunit D).
GN ATPD OR VATD.
GN Thermus thermophilus.
OS Thermus thermophilus.
OC Bacteria; Thermus/Delinococcus group; Deinococci; Thermales;
OC Thermaceae; Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB8 / ATCC 27634;
RX MEDLINE=20250964; PubMed=10788522;
RA Yokoyama K., Onkuma S., Taguchi H., Yasunaga T., Wakabayashi T.,
RA Yoshida M.;
RT "V-type H⁺-ATPase/synthase from a thermophilic eubacterium, Thermus
RT thermophilus. Subunit structure and operon.";
RL J. Biol. Chem. 275:13955-13961(2000).
CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SIMILARITY: BELONGS TO THE V-ATPASE D SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; D63799; BAA33198.1; ALT_INIT.
DR InterPro; IPR002699; ATPsynT_Dsub.
DR Pfam; PF01813; ATP-synt_D; 1.
DR ProDom; PD004122; ATPsynT_Dsub; 1.
DR TIGRfams; TIGR00309; V_ATPase_subD; 1.
KW Hydrolyase; ATP synthetase; Hydrogen ion transport.
SQ SEQUENCE 223 AA; 24677 MW; B66AE00DFE0B08B CRC64;

Query Match 0.8%; Score 7; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 179 KVPRLKA 185
 DB 100 KVPRLKA 106

RESULT 20

GPH_VIBCH STANDARD: PRT: 226 AA.
 ID GPH_VIBCH
 AC 09KNV6;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Phosphoglycolate phosphatase (EC 3.1.3.18) (PGP).
 GN GPH OR VC2624.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae.";
 RL Nature 406:477-483(2000).
 CC -1- CATALYTIC ACTIVITY: 2-phosphoglycolate + H(2)O = glycolate +
 CC phosphate.
 CC -1- SIMILARITY: BELONGS TO THE CBBY/CBBZ/GPH/YIEH FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AEO04329; AAF95765.1; -;
 DR TIGR: VC2624; -;
 DR InterPro: IPR001454; Hlgase/hydrlase.
 DR Pfam: PF00702; Hydrolase; 1.
 KM Carboxylate metabolism; Hydrolase; Complete proteome.
 SQ SEQUENCE 226 AA; 24560 MW; 83F95B10DBF53CB3 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 226;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 241 AARAVGY 247
 DB 29 AARAVGY 35

RESULT 21

COX2_ASCSU STANDARD: PRT: 232 AA.
 ID COX2_ASCSU
 AC P24882;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1).
 GN COII.
 OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
 OC Mitochondrion
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
 OC Ascarididae; Ascaris.
 OX NCBI_TaxID=6253;
 RN [1]
 RP SEQUENCE FROM N.A.

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=body wall muscle, and Egg;
 RX MEDLINE=92201635; PubMed=1551572;
 RA Okimoto R., Macfarlane J.L., Clary D.O., Wolstenholme D.R.;
 RT "The mitochondrial genomes of two nematodes, Caenorhabditis elegans
 and Ascaris suum.";
 RL Genetics 130:471-498(1992).

CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 2
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
 CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
 CC -1- CATALYTIC ACTIVITY: 4 ferrioxochrome c + O(2) = 4 ferrioxochrome
 CC c + 2 H(2)O.
 CC -1- COFACTOR: Copper A.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X54253; CAA38172.1; -;
 DR PIR: S26023; S26023.
 DR InterPro: IPR001505; Copper_CuA.
 DR InterPro: IPR002429; Cyl_C-ox_2.
 DR Pfam: PF00116; COX2; 1.
 DR Pfam: PF02790; COX2_TM; 1.
 DR ProDom: PD000131; Copper_CuA; 1.
 DR PROSITE: PS00078; COX2; 1.
 KM Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
 KM Electron transport; Respiratory chain.
 FT DOMAIN 1 30 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
 FT TRANSMM 53 69 POTENTIAL.
 FT DOMAIN 31 52 MITOCHONDRIAL MATRIX (POTENTIAL).
 FT TRANSMM 70 89 POTENTIAL.
 FT DOMAIN 90 232 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
 FT METAL 164 164 COPPER A (PROBABLE).
 FT METAL 199 199 COPPER A (PROBABLE).
 FT METAL 203 203 COPPER A (PROBABLE).
 FT METAL 207 207 COPPER A (PROBABLE).
 SQ SEQUENCE 232 AA; 26517 MW; 58F9B683FF9DA4C5 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 232;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 SFMPVAL 17
 DB 208 SFMPVAL 214

RESULT 22

RSTA_ECOLI STANDARD: PRT: 242 AA.
 ID RSTA_ECOLI
 AC P52106;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcriptional regulatory protein rsta.
 GN RSTA OR URPT OR B1608.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-K12;
RA Kuempel P.L.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blatner F.R., Plunkett G., Ilt, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97251357; PubMed-9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubraman S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [4]
RP MAPPING.
RX MEDLINE-92356829; PubMed-1495392;
RA Roedelstein B.A., Kuempel P.L.;
RT "In vivo characterization of tus gene expression in Escherichia
coli.";
RL Mol. Microbiol. 6:1655-1661(1992).
CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM RSTB/RSTA.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- PTM: PHOSPHORYLATED BY RSTB (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U41101; AAA82081.1; -;
DR EMBL: AE000256; AAC74680.1; -;
DR EMBL: D90803; BAA15346.1; -;
DR EMBL: D90804; BAA15356.1; -;
DR HSP: P03025; IOPC.
DR Ecogene: EG31390; rsta.
DR InterPro: IPR001789; Response_reg.
DR InterPro: IPR001867; Trans_reg.C.
DR Pfam: PF00072; response_reg.1.
DR Pfam: PF00486; trans_reg.C.1.
DR ProDom: PD000039; Response_reg.1.
DR ProDom: PD000329; Trans_reg.C.1.
DR SMART: SM00448; RRC.1.
DR PROSITE: PS50110; RESPONSE_REGULATORY.1.
DR Sensory transduction; Phosphorylation; Transcription regulation;
KW DNA-binding; Complete proteome.
FT MOD_RES 55 119 RESPONSE REGULATORY.
FT MOD_RES 55 119 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 242 AA; 27048 MW; 8B5A3068829811B6 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 242;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 334 LDRSVYV 340
|||||

Db 201 LDRSVYV 207
RESULT 23
CLDI_HUMAN
ID CLDI_HUMAN STANDARD; PRT: 261 AA.
AC P56856; Q96PH4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Claudin-18.
CN CLDN18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE-21470315; PubMed-11585919;
RA Nishimura T., Nagashima K., Ward J.M., Minoo P., Zimonjic D.B.,
RA Popescu N.C., Kimura S.;
RT "Claudin-18, a novel downstream target gene for the T/EBP/NKX2.1
homeodomain transcription factor, encodes lung- and stomach-specific
isoforms through alternative splicing.";
RL Mol. Cell. Biol. 21:7380-7390(2001).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: A1 (shown here) and A2; are
CC produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Concentrated at the cell-cell borders of
CC epithelial cells.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF221069; AAF26448.1; -;
DR EMBL: AF349452; AAL15637.1; -;
DR Genew: HGNC:2039; CLDN18.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR004031; PMP22.Claudin.
DR Pfam: PF00822; PMP22.Claudin; 1.
DR PROSITE: PS01346; CLAUDIN.1.
KW Tight junction; Transmembrane; Alternative splicing.
FT TRANSMEM 7 27
FT TRANSMEM 81 101
FT TRANSMEM 123 143
FT TRANSMEM 175 195
FT VARSPLIC 1 69
FT
FT MSTRVQVAFLLSIIGACIAATGDMSTODLYDNPVT
FT SFVYEGILMRSVCROSSGFETCRPYFTI -> MAVTACGL
FT GFVYELIGIIGIAATGDMSTODLYDNPVTAFFVNGGLW
FT RSCVRESSGFTCCRGYFTL (IN ISOFORM A2).
SQ SEQUENCE 261 AA; 27856 MW; 4362B590D3C8B387 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 261;

Best Local Similarity 100.0%; Pred. No. 55;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 404 LQAVRAL 410
|||||

Db 76 LQAVRAL 82

RESULT 24

CLDI_MOUSE STANDARD; PRT: 264 AA.

AC P56857; Q91ZL1; Q91Z20; Q91Z9;

DT 30-MAY-2000 (Rel. 39, Created)

```

DT 30-MAY-2000 (Rel. 39, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE CLAUDIN-18.
GN CLDN18.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=21470315; PubMed=11585919;
RA Nishikawa T., Nagashima K., Ward J.M., Munoo P., Zimonjic D.B.,
RA Popescu N.C., Kimura S.;
RT "Claudin-18, a novel downstream target gene for the T/EBV/NAK2.1
RT homeodomain transcription factor, encodes lung- and stomach-specific
RT isoforms through alternative splicing."
RL Mol. Cell. Biol. 21:7380-7390(2001).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; A1.1 (shown here), A1.2, A2.1
CC and A2.2: are produced by alternative splicing.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF221068; AAF26447.1; -
DR EMBL; AF349450; AAL15635.1; -
DR EMBL; AF349451; AAL15636.1; -
DR EMBL; AF349453; AAL15638.1; -
DR MGI; MGI:1929209; Clnd18.
DR InterPro: IPR001832; Claudin.
DR Pfam: PF00822; PMP22.Claudin.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane; Alternative splicing.
SQ
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT VARSPIC 1 69 MATTCOVYGLLSLLGIACIAATGMDMSTODLYDNVPT
FT AVFOYEGLMRSCVOSSGFTECRPYFI -> MGVTAQGL
FT GFVVSLLGFAGIIATCNDMDSTODLYNNPTAVFNVOGLW
FT RSCVRESGFECEGRTFL (IN ISOFORM A2.1 AND
FT ISOFORM A2.2).
FT N -> K (IN ISOFORM A1.2 AND ISOFORM
FT A2.2).
FT VARSPIC 208 208 MISSING (IN ISOFORM A1.2 AND ISOFORM
FT VARSPIC 209 264 MISSING (IN ISOFORM A1.2 AND ISOFORM
FT A2.2).
SQ SEQUENCE 264 AA; 28122 MW; 3CA0D441C4705653 CRC64;
Query Match 0.8%; Score 7; DB 1; Length 264;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 404 LQAVRAL 410
DB 76 LQAVRAL 82

```

```

RESULT 25
PURR_LACIA STANDARD; PRT; 271 AA.
AC 053065;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pur operon repressor.

```

```

GN PURR OR LL2259.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis), and
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360, 1359;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=L.l.lactis; STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Mincker P., Manger S., Jallion O., Malarme K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RT Genome Res. 11:731-753(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=L.l.cremoris; STRAIN=MGI1363;
RA Kilstrup M.;
RT Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CONTROLS TRANSCRIPTION OF THE PUR OPERON FOR PURINE
CC BIOSYNTHETIC GENES (BY similarity).
CC -1- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE
CC PHOSPHORIBOSYLTRANSFERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AJ222642; CAA10902.1; -
DR EMBL; AE006455; AAK06357.1; -
DR InterPro: IPR000836; PRTtransferase.
DR InterPro: IPR002375; Pr/PY-TP-Transf.
DR Pfam: PF00156; Pribosyltran; 1.
DR PROSITE: PS00103; PUR_PYR_PR_TRANSFER; FALSE NEG.
KW DNA-binding; Transcription regulation; Repressor; Complete proteome.
SQ SEQUENCE 271 AA; 30361 MW; 04614AA24E1C4BCD CRC64;

```

```

Query Match 0.8%; Score 7; DB 1; Length 271;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 722 LLAGYAV 728
DB 223 LLAGYAV 229

```

```

RESULT 26
LIGE_PSEPA STANDARD; PRT; 280 AA.
ID LIGE_PSEPA
AC P27457;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Beta-etherase (Beta-aryl ether cleaving enzyme).
GN LIGE.
OS Pseudomonas paucimobilis (Sphingomonas paucimobilis).
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Sphingomonas.
OX NCBI_TaxID=13689;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC STRAIN=SYK-6;
RX MEDLINE=92078104; PubMed=1744051;
RA Masai E., Katayama Y., Kawai S., Nishikawa S., Yamasaki M.,
RA Morohoshi N.;
RT "Cloning and sequencing of the gene for a Pseudomonas paucimobilis
RT enzyme that cleaves beta-aryl ether."
RT J. Bacteriol. 173:7950-7955(1991).
CC -1- FUNCTION: ABLE TO DEGRADE VARIOUS DIMERIC LIGNIN COMPOUNDS.

```

RA	Azavedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA	Borries R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
RA	Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA	Choi S.K., Codani J.J., Conerton I.F., Cunningham N.J., Daniel R.A.,
RA	Denizot F., Devine K.M., Dusterhoft A., Ehlich S.D., Emerson P.T.,
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Gallero N.,
RA	Ghim S.Y., Glaser P., Goffeau A., Goldlight E.J., Grandi G.,
RA	Guseppl G., Guy B.J., Haga K., Haleh J., Harwood C.R., Hehaut A.,
RA	Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA	Kobayashi Y., Koester P., Konigstein G., Krogh S., Kumano M.,
RA	Kunita K., Lapidos A., Lardinois S., Lauber J., Lazarevic V.,
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA	Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Nodack M.,
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA	Parro V., Pohl T.M., Portetelle D., Porvolik S., Prescott A.M.,
RA	Prescott E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA	Seikunghi J., Sekowska A., Serot S.J., Serrot P., Shin B.S., Soldo B.,
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA	Takenuchi M., Tamakoshi A., Tanaka T., Tarpstra P., Tognoni A.,
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA	Viati A., Wambut R., Wedler E., Wedler H., Weltzenegger T.,
RA	Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA	Yoshida K., Yoshikawa H.F., Zumsstein E., Yoshikawa H., Danchin A.,
RT	"The complete genome sequence of the Gram-positive bacterium <i>Bacillus</i>
RT	<i>subtilis</i> ".
RL	Nature 390:249-256(1997).
CC	-i- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC	-i- SIMILARITY: TO B.SUBTILIS YPIC AND YQPU.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL, D16311; BAA03818.1; -
DR	EMBL, Y09476; CAA70630.1; -
DR	EMBL, Z79580; CAB01834.1; -
DR	EMBL, Z99109; CAB12952.1; -
DR	PIR: A49921; A49921.
DR	Subtilist: BG10850; Ylft.
DR	InterPro: IPR003740; DUF161.
DR	Pfam: PF02588; DUF161, 1.
KW	Hypothetical protein; Transmembrane; Complete proteome.
FT	TRANSMEM 9 29 POTENTIAL.
FT	TRANSMEM 54 74 POTENTIAL.
FT	TRANSMEM 80 100 POTENTIAL.
FT	TRANSMEM 151 171 POTENTIAL.
FT	CONFLICT 121 121 V -> D (IN REF. 1).
FT	CONFLICT 226 226 V -> L (IN REF. 1).
FT	CONFLICT 278 278 R -> T (IN REF. 1).
SO	SEQUENCE 280 AA; 30535 MW; 4BF98B1288DA59E7 CNC64;
QY	70 LNAAGLN 76
Db	18 LNAAGLN 24
RESULT 28	
THTR_HUMAN	
ID THTR_HUMAN	STANDARD; PRT; 296 AA.
AC Q16762;	
DT 15-JUL-1998 (Rel. 36, Created)	

DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Thiosulfate sulfotransferase (EC 2.8.1.1) (Rhodanese).
 GN TST.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal Liver;
 RX MEDLINE=97223396; PubMed=9070219;
 RA Alta N., Ishii K., Akamatsu Y., Ogasawara Y., Tanabe S.;
 RT Cloning and expression of human liver rhodanese cDNA.
 RL Biochem. Biophys. Res. Commun. 231:56-60(1997).
 RP [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
 RA Clamp M., Slink L.J., Alnscough R., Almeida J.P., Habbage A.K.,
 RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
 RA Burdill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole G.G., Collier R.E., Connor R.,
 RA Conroy D., Corry N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Graffham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Lait D.G., Langford C.F., Leverisha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashieghl-Mohammadi M., Matthews L.H., Mccann O.T.,
 RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.,
 RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Mioshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyma S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.T.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
 RA Wang O., Wang Y., Wang Z., White J., Willingham D., Wu R., Yao Z.,
 RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozeresky P., Rohlfing T.,
 RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
 RA Emanuel B.S., Shaikh T., Kurahashi H., Saita S., Budarf M.L.,
 RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
 RA Kim U.-J., Shizuya H., Simon M.I., Dumanaki J.P., Peyraud M., Kedra D.,
 RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
 RA Wilkinson P., Podeweltch A., Hartman K., Hu X., Khan A.S., Lane L.,
 RA Tlahun Y., Wright H.;
 RT "The DNA sequence of human chromosome 22."
 RL Nature 402:489-495(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: FORMATION OF IRON-SULFUR COMPLEXES. CYANIDE
 CC DETOXIFICATION OR MODIFICATION OF SULFUR-CONTAINING ENZYMES.
 CC OTHER THIOL COMPOUNDS, BESIDES CYANIDE, CAN ACT AS SULFUR ION
 CC ACCEPTORS. ALSO HAS WEAK MERCAPTOPROPYLATE SULFOTRANSFERASE (MST)
 CC ACTIVITY (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.
 CC -1- SUBUNIT: MONOMER.

CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -1- DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR
 CC CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,
 CC THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D87292; BAA13327.1; -;
 CC EMBL: 273420; CAA97762.1; -;
 CC EMBL: BC010148; AAH0148.1; -;
 CC HSSP: P00586; IRHS.
 CC Genew: HGNC:12388; TST.
 CC MIM: 180370; -;
 CC InterPro: IPR001763; Rhodanese-like.
 CC InterPro: IPR001307; Rhodanese.
 CC Pfam: PF00581; Rhodanese; 2.
 CC SMART: SM00450; RHOD; 2.
 CC PROSITE: PS00380; RHODANES_1; 1.
 CC PROSITE: PS00683; RHODANES_2; 1.
 CC KW Transferase; Mitochondrion.
 CC FT INIT MET 0
 CC FT DOMAIN 1 142 BY SIMILARITY.
 CC FT DOMAIN 143 158 A DOMAIN.
 CC FT DOMAIN 159 296 HINGE.
 CC FT ACT_SITE 186 196 B DOMAIN.
 CC FT ACT_SITE 247 247 MAY PLAY A ROLE IN SUBSTRATE BINDING (BY
 CC ACT_SITE 248 248 SIMILARITY).
 CC ACT_SITE 249 249 SUBSTRATE (THIOSULFATE) BINDING
 CC (BY SIMILARITY).
 CC SEQUENCE 296 AA; 33297 MW; 872C52008AE8DC5B CRC64;
 CC
 CC Query Match 0.8%; Score 7; DB 1; Length 296;
 CC Best Local Similarity 100.0%; Pred. No. 62;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 15 VALAAVL 21
 CC 1111111
 CC Db 256 VALAAVL 262
 CC
 CC RESULT 29
 CC THTR_MOUSE
 CC ID THTR_MOUSE STANDARD; PRT; 296 AA.
 CC AC P52196;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
 CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
 CC DE Thiosulfate sulfotransferase (EC 2.8.1.1) (Rhodanese).
 CC GN TST.
 CC OS Mus musculus (Mouse).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC OX NCBI_TaxID=10090;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=BALB/C; TISSUE=Liver;
 CC RX MEDLINE=96074596; PubMed=7488186;
 CC RA Dooley T.P., Nair S.K., Garcia R.E., Courtney B.C.;
 CC "Mouse rhodanese gene (Tst): cDNA cloning, sequencing, and
 CC recombinant protein expression."
 CC RT Biochem. Biophys. Res. Commun. 216:1101-1109(1995).
 CC -1- FUNCTION: FORMATION OF IRON-SULFUR COMPLEXES AND CYANIDE
 CC DETOXIFICATION.
 CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.
 CC -1- SUBUNIT: MONOMER.


```
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -1- TISSUE SPECIFICITY: FOUND IN NUMEROUS TISSUES.
CC -1- DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR
CC CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,
CC THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.
CC -1- SIMILARITY: BELONGS TO THE RHODANSE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U35741; AAC52342.1; -.
CC HSSP: P00586; 1RHS.
CC SWISS-2DPAGE: P52196; MOUSE.
CC MGI: 98852; Tst.
CC InterPro: IPR001763; Rhodanese-like.
CC InterPro: IPR001307; Rhodanese.
CC Pfam: PF00581; Rhodanese; 2.
CC SMART: SM00450; RHOD. 2.
CC PROSITE: PS00380; RHODANSE_1; 1.
CC PROSITE: PS00683; RHODANSE_2; 1.
CC Transferase; Mitochondrion.
CC INIT_MET 0
CC DOMAIN 1 142 A DOMAIN.
CC FT DOMAIN 143 158 HINGE.
CC FT DOMAIN 159 296 B DOMAIN.
CC FT ACT_SITE 186 186 MAY PLAY A ROLE IN SUBSTRATE BINDING (BY
CC ACT_SITE 247 247 SIMILARITY).
CC FT ACT_SITE 248 248 BY SIMILARITY.
CC FT ACT_SITE 249 249 SUBSTRATE (THIOSULFATE) BINDING
CC (BY SIMILARITY).
CC SEQUENCE 296 AA; 33334 MW; 82089D80F9AE55A CRC64;
Query Match 0.8%; Score 7; DB 1; Length 296;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 15 VALAAYL 21
|111111|
Db 256 VALAAYL 262

RESULT 30
34KD_MYCTU STANDARD; PRT; 303 AA.
ID 34KD_MYCTU
AC P71556;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 34 kDa antigenic protein homolog.
GN RV0954 OR MT0981 OR MTCY10D7.20C.
OS Mycobacterium tuberculosis.
CC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CC NCBI_TaxID=1773;
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
```

```
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: STRONG, TO M.PARATUBERCULOSIS 34 kDa ANTIGENIC
CC PROTEIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: Z79700; CAB01996.1; -.
CC DR EMBL: AE006983; AKK45229.1; -.
CC DR TIGR: MT0981; -.
CC TubercuList; RV0954; -.
CC FT Transmembrane; Complete proteome.
CC FT TRANSMEM 42 62 POTENTIAL.
CC FT TRANSMEM 77 97 POTENTIAL.
CC FT TRANSMEM 102 122 POTENTIAL.
CC FT TRANSMEM 134 154 POTENTIAL.
CC SEQUENCE 303 AA; 30204 MW; 4FE18A077FDFDB5 CRC64;
Query Match 0.8%; Score 7; DB 1; Length 303;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 721 ALLAGVA 727
|111111|
Db 88 ALLAGVA 94

RESULT 31
OYE6_HUMAN STANDARD; PRT; 313 AA.
ID OYE6_HUMAN
AC Q96RD3;
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Olfactory receptor 52B6.
GN OR52B6.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CC NCBI_TaxID=9606;
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 70-287 FROM N.A.
RA Fuchs T., Malecova B., Linhart C., Sharan R., Khen M., Herwig R.,
RA Shmulevich D., Elkon R., Steinfelt M., O'Brien J.K., Radelof U.,
RA Lehrach H., Olander Z., Glusman G., Lancet D., Shamir R.;
RT "DEFOG: a practical scheme for deciphering families of genes.";
CC Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AF399504; AAK94989.1; -
 InterPro: IPR000276; GPCR_Rhodopsin.
 DR Pfam: PF00001; 7tm.1; 1.
 DR PROSITE: PS00237; G_PROTEIN_RECPT_FL1; 1.
 DR PROSITE: PS50262; G_PROTEIN_RECPT_FL2; 1.
 KW G-protein coupled receptor; Transmembrane; glycoprotein;
 MultiGene family; Olfaction.
 FT DOMAIN 1 27 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 28 51 1 (POTENTIAL).
 FT DOMAIN 52 59 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 60 81 2 (POTENTIAL).
 FT DOMAIN 82 102 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 103 122 3 (POTENTIAL).
 FT DOMAIN 123 141 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 142 160 4 (POTENTIAL).
 FT DOMAIN 161 197 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 198 220 5 (POTENTIAL).
 FT DOMAIN 221 237 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 238 260 6 (POTENTIAL).
 FT DOMAIN 261 275 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 276 295 7 (POTENTIAL).
 FT DOMAIN 296 313 CYTOPLASMIC (POTENTIAL).
 FT CAROHPD 5 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 313 AA; 35523 MW; 6FC174DBDC20BB3 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 313;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 VVVPPTL 157
 |||||
 Db 282 VVVPPTL 288

RESULT 32
 MRAM_FUSUN STANDARD; PRT; 314 AA.
 AC Q8R6F5;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE S-adenosyl-methyltransferase mram (EC 2.1.1.-).
 GN MRAM OR FNI711.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 CC Bacteria; Fusobacteria; Fusobacterium.
 OX NCBI_TaxID=76856;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RX MEDLINE=21886394; PubMed=11889109;
 RA Kapural V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Vastava O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
 RA Fongsteil M., Kyrpides N., Overbeek R.;
 RT Genome sequence and analysis of the oral bacterium Fusobacterium
 nucleatum strain ATCC 25586. ;
 RL J. Bacteriol. 184:2005-2018(2002).
 CC -1- FUNCTION: Exhibits a S-adenosyl-dependent methyltransferase
 activity (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE MRAM FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL: AE010475; AAL93826.1; -
 KW Transferase; Methyltransferase; Complete proteome.
 SQ SEQUENCE 314 AA; 35782 MW; AB1FD9DD3D269136 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 314;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 471 DGLIMDI 477
 |||||
 Db 100 DGLIMDI 106

RESULT 33
 MOCA_RHIME STANDARD; PRT; 317 AA.
 AC P49307;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Rhizopine catabolism protein moca (EC 1.-.-.-).
 GN MOCA.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=L5-30;
 RX MEDLINE=95147842; PubMed=7845353;
 RA Rossbach S., Kulpa D.A., Rossbach U., de Bruijn F.J.;
 RT "Molecular and genetic characterization of the rhizopine catabolism
 (mocaBrc) genes of Rhizobium meliloti L5-30." ;
 RL Mol. Gen. Genet. 245:11-24(1994).
 CC -1- INVOLVED IN RHIZOPINE (L-3-O-METHYL-SCYLLA-1NOSAMINE) CATABOLISM.
 CC -1- SIMILARITY: BELONGS TO THE GFO/IDH/MOCA FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL: X78503; CAA55269.1; -
 InterPro: IPR000683; GFO_IDH_MOCA.
 DR InterPro: IPR004104; GFO_IDH_MOCA_C.
 DR Pfam: PF01408; GFO_IDH_MOCA; 1.
 KW Oxidoreductase; NAD.
 SQ SEQUENCE 317 AA; 33725 MW; 06CD96EC77276652 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 317;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 752 GSSGLVS 758
 |||||
 Db 220 GSSGLVS 226

RESULT 34
 ISPB_ECOLI STANDARD; PRT; 323 AA.
 AC P19641;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Octaprenyl-diphosphate synthase (EC 2.5.1.-) (Octaprenyl pyrophosphate

DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Iron-utilization periplasmic protein precursor (Major ferric iron
 binding protein) (FBP) (Iron-regulated 40 kDa protein) (MIRP) (Fe(+3)-
 binding protein).
 GN HTTA OR FBP OR H10097.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 Haemophilus.
 NCBI_TaxID:727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fieleschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
 Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 Venter J.C.;
 RA *Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.;
 RL Science 269:496-512(1995).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-NTHI TN106;
 RX MEDLINE=95012644; PubMed=7927717;
 RA Sanders J.D., Cope L.D., Hansen E.J.;
 RT *Identification of a locus involved in the utilization of iron by
 Haemophilus influenzae.;
 RL Infect. Immun. 62:4515-4525(1994).
 RN [3]
 RP SEQUENCE OF 24-45.
 RC STRAIN-EAGAN / 8358;
 RX MEDLINE=92210482; PubMed=1556062;
 RA Hartness R.E., Chong P., Klein M.H.;
 RT *Identification of two iron-repressed periplasmic proteins in
 Haemophilus influenzae.;
 RL J. Bacteriol. 174:2425-2430(1992).
 RN [4]
 RP SEQUENCE OF 24-31.
 RC MEDLINE=20137488; PubMed=10675023;
 RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
 Gray C., Fountoulakis M.;
 RT *Two-dimensional map of the proteome of Haemophilus influenzae.;
 RL Electrophoresis 21:411-429(2000).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 24-332.
 RX MEDLINE=98025060; PubMed=9360608;
 RA Bruns C.M., Nowalk A.J., Arvai A.S., McTigue M.A., Vaughan K.G.,
 Mietner T.A., McGee D.E.;
 RT *Structure of Haemophilus influenzae Fe(+3)-binding protein reveals
 convergent evolution within a superfamily.;
 RL Nat. Struct. Biol. 4:919-924(1997).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS) OF 24-332.
 RX MEDLINE=21614394; PubMed=11747438;
 RA Bruns C.M., Anderson D.S., Vaughan K.G., Williams P.A., Nowalk A.J.,
 McGee D.E., Mietner T.A.;
 RT *Crystallographic and biochemical analyses of the metal-free
 Haemophilus influenzae Fe(+3)-binding protein.;
 RL Biochemistry 40:15631-15637(2001).
 CC -1- FUNCTION: THIS PROTEIN MAY BE A CENTRAL COMPONENT IN THE IRON-
 ACQUISITION SYSTEM.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- INDUCTION: BY IRON DEPRIVATION.
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
 PROTEIN FAMILY 1.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A

CC ----- FRAMESHIFT IN POSITION 14.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U32695; AAC21773.1; ALT_FRAME.
 DR EMBL: S72674; AAB32110.1; -.
 DR PIR: A41833; A41833.
 DR PDB: 1MRP; 28-JAN-98.
 DR PDB: 1D9V; 17-NOV-99.
 DR TIGR: H10097; -.
 DR InterPro: IPR000567; SBP_bac_1.
 DR Pfam: PF01547; SBP_BACTERIAL_1; 1.
 DR PROSITE: PS01037; SBP_BACTERIAL_1; 1.
 KW Transports; Iron transport; Periplasmic; Signal; 3D-structure;
 KW Complete proteome.
 FT SIGNAL 1 23
 FT CHAIN 24 332
 FT VARIANT 24 24
 FT VARIANT 37 37
 FT VARIANT 37 37
 FT VARIANT 37 37
 FT VARIANT 59 59
 FT VARIANT 103 103
 FT VARIANT 119 119
 FT VARIANT 284 284
 FT VARIANT 323 323
 SQ SEQUENCE 332 AA; 36193 MW; 4E6549PD76262CA6 CRC64;
 Query Match 0.8%; Score 7; DB 1; Length 332;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 238 ALVAAA 244
 Db 280 ALVAAA 286
 ID YPFG_ECOLI STANDARD; PRT; 347 AA.
 AC P76559;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein yPFG precursor.
 GN YPFG OR B2466.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 Mau B., Shao Y.;
 RT *The complete genome sequence of Escherichia coli K-12.;
 RL Science 277:1234-1238(2000).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC -----
DR EMBL: AE00333; AAC75519.1; -.
DR Ecocore; EC14194; ypfG.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 347
SQ SEQUENCE 347 AA; 38746 MW; 1E7566CCE3B066FF CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 347;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 TP1SLSE 122
Db 193 TP1SLSE 199

RESULT 38
GP21_HUMAN
ID GP21_HUMAN STANDARD; PRT; 349 AA.
AC Q99679;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable G protein-coupled receptor GPR21.
GN GPR21.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-972579; PubMed-9073069;
RA O'Dowd B.F., Nguyen T., Jung B.P., Marchese A., Cheng R.,
RA Heng H.H.O., Kolakowski L.F. Jr., Lynch K.R., George S.R.;
RT "Cloning and chromosomal mapping of four putative human
RT G-protein-coupled receptor genes.";
RL Gene 187:75-81(1997).
CC -1- FUNCTION: ORPHAN RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: NOT DETECTED IN THE BRAIN REGIONS THALAMUS,
CC PUTAMEN, CAUDATE, FRONTAL CORTEX, PONS, HYPOTHALAMUS, HIPPOCAMPUS.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U66580; AAC51303.1; -.
DR Genew; HGNC:4476; GPR21.
DR MIM: 601909;
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm_1. 1.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 32
FT TRANSMEM 33 53
FT DOMAIN 54 75
FT TRANSMEM 76 96
FT DOMAIN 97 104
FT TRANSMEM 105 125
FT DOMAIN 126 147
FT TRANSMEM 148 168
FT DOMAIN 169 191
FT TRANSMEM 192 212
FT DOMAIN 213 252
FT TRANSMEM 253 273
FT DOMAIN 274 283
EXTRACELLULAR (POTENTIAL).

```

```

FT TRANSMEM 284 304
FT DOMAIN 305 349
FT CARBOHYD 2 2
FT CARBOHYD 8 8
SQ SEQUENCE 349 AA; 39515 MW; 5369DFC16E9EEC4 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 349;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 SSOSGET 196
Db 233 SSOSGET 239

RESULT 39
HRCA_CAUCR
ID HRCA_CAUCR STANDARD; PRT; 358 AA.
AC P54305;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heat-inducible transcription repressor hrca.
GN HRCA OR CC0153.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CB15N / NA1000;
RA MEDLINE-96178944; PubMed-8606155;
RA Roberts R.C., Toochinda C., Avedissian M., Baldini R.L.,
RA Gomes S.L., Shapiro L.;
RT "Identification of a Caulobacter crescentus operon encoding hrca,
RT involved in negatively regulating heat-inducible transcription, and
RT the chaperrone gene grpe.";
RL J. Bacteriol. 178:1829-1841(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 19089 / CB15;
RA MEDLINE-21173698; PubMed-11259647;
RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
RA Klonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback J., Tran K., Wolf A., Yamathavan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -1- FUNCTION: NEGATIVE REGULATOR OF CLASS I HEAT SHOCK GENES (GRPE-
CC DNAR-DNAJ AND GROELS OPERONS). PREVENTS HEAT-SHOCK INDUCTION OF
CC THESE OPERONS.
CC
CC -1- SIMILARITY: BELONGS TO THE HRCA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U33324; AAB01515.1; -.
DR EMBL: AE005689; AAK22140.1; -.
DR TIGR: CC0153;
DR InterPro: IPR002571; Hrca.
DR Pfam: PF01628; Hrca; 1.
DR TIGRFRAMS: TIGR00331; hrca; 1.
KW Transcription regulation; Repressor; Heat shock; Complete proteome.
FT CONFLICT 276 279
FT SEQUENCE 358 AA; 38265 MW; 38D7D61D94D3B13C CRC64;

```

Query Match 0.8%; Score 7; DB 1; Length 358;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 838 GGOVLAV 844
 DB 158 GGOVLAV 164

RESULT 40
 PCE_TACTR
 ID PCE_TACTR STANDARD; PRT: 375 AA.
 AC P21902;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Proclotting enzyme precursor (EC 3.4.21.86).
 OS Tachypleus tridentatus (Japanese horseshoe crab).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
 OC Limulidae; Tachypleus.
 OX NCBI_TaxID=6853;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RC TISSUE=Hemocyte;
 RX MEDLINE=91093088; PubMed=2266134;
 RA Muta T., Hashimoto R., Miyata T., Nishimura H., Toh Y., Iwanaga S.;
 RT "Proclotting enzyme from horseshoe crab hemocytes. cDNA cloning,
 disulfide locations, and subcellular localization.";
 RL J. Biol. Chem. 265:22426-22433(1990).
 CC -1- FUNCTION: THIS ENZYME IS CLOSELY ASSOCIATED WITH AN ENDOTOXIN-
 SENSITIVE HEMOLYMPH COAGULATION SYSTEM IN LIMULUS. ITS ACTIVE FORM
 CATALYZES THE CONVERSION OF COAGULOGEN TO INSOLUBLE COAGULIN GEL.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of 18-Arg-I- and 47-Arg-I-
 bonds in coagulogen to form coagulin and fragments.
 CC -1- SUBUNIT: UPON ACTIVATION BY FACTOR B, IT IS CONVERTED TO A TWO-
 CHAIN ACTIVE FORM COMPOSED OF A LIGHT AND A HEAVY CHAIN.
 CC -1- PTM: CONTAINS SIX O-LINKED CARBOHYDRATE CHAINS IN THE N-TERMINAL
 LIGHT CHAIN GENERATED AFTER ACTIVATION.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M58366; AAA30094.1; -;
 DR PIR: A23689; A23689.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.221; -;
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR Pfam: PF00089; Trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYD_SPC; 1.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KW Hydrolyase; Serine protease; Zymogen; Signal; Hemolymph clotting;
 KW Glycoprotein.
 FT SIGNAL 1 17
 FT PROPEP 18 29 OR 25 (POTENTIAL).
 FT CHAIN 30 127 PROCLOTTING ENZYME LIGHT CHAIN.
 FT CHAIN 128 375 PROCLOTTING ENZYME HEAVY CHAIN.
 FT DOMAIN 128 375 SERINE PROTEASE.
 FT MOD_RES 30 30 PYRROLIDONE CARBOXYLIC ACID.
 FT ACT_SITE 172 172 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 228 228 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 326 326 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 40 83

FT DISULFID 50 73
 FT DISULFID 56 84
 FT DISULFID 118 248 INTERCHAIN.
 FT DISULFID 157 173
 FT DISULFID 295 311
 FT DISULFID 322 351
 FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 235 235 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .).
 SQ SEQUENCE 375 AA; 41591 MW; 94209E514DFCF8FB CRC64;

Query Match 0.8%; Score 7; DB 1; Length 375;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 698 TOAPPET 704
 DB 94 TOAPPET 100

Search completed: April 28, 2003, 16:33:39
 Job time : 66 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:30:40 ; Search time 101 Seconds
(without alignments)
1874.825 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 919
Sequence: 1 MSKPYLFANRSMFVALAAY.....TGKKEGNDPIKLFITGTFP 919

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 1000 summaries

Database :

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *
15: SP virus: *
16: SP bacteriophage: *
17: SP archaea: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	1.4	578	16	082B95
2	11	1.2	577	16	08XB95
3	10	1.1	579	16	0910U1
4	9	1.0	260	16	0849S1
5	9	1.0	262	2	09R802
6	9	1.0	586	16	09CK26
7	9	1.0	1085	2	09RPH7
8	8	0.9	174	16	P95086
9	8	0.9	235	16	098UV3
10	8	0.9	319	16	096HK9
11	8	0.9	331	16	0927W8
12	8	0.9	331	16	08Y4C5
13	8	0.9	340	16	09ZBL4
14	8	0.9	379	2	P72239
15	8	0.9	383	16	09KYD1
16	8	0.9	452	12	08V5A8

17	8	0.9	469	16	086499	086499 streptomyces
18	8	0.9	508	5	076799	076799 diatomella
19	8	0.9	508	17	08T139	08T139 methanobacterium
20	8	0.9	582	16	09KP31	09KP31 vibrio cholerae
21	8	0.9	731	16	08ZS53	08ZS53 anabaena sp
22	8	0.9	1047	17	09HK17	09HK17 thermoplasma
23	8	0.9	1373	5	09VKR2	09VKR2 diatomella
24	8	0.9	1417	10	048697	048697 arachidonic acid
25	8	0.8	33	2	045283	045283 bacillus
26	7	0.8	62	9	08W6C9	08W6C9 vibrio phage
27	7	0.8	66	10	08W542	08W542 retama raet
28	7	0.8	66	16	09PD36	09PD36 xylella fastidiosa
29	7	0.8	72	16	P74009	P74009 synecocyst
30	7	0.8	82	16	08XGJ5	08XGJ5 salmoneella
31	7	0.8	84	17	08TRE1	08TRE1 methanobacterium
32	7	0.8	85	13	09PVI8	09PVI8 gallus galli
33	7	0.8	87	16	09CLV3	09CLV3 pasteurella
34	7	0.8	88	16	08Z102	08Z102 yersinia pestis
35	7	0.8	91	16	09J5V0	09J5V0 neisseria meningitidis
36	7	0.8	95	12	098552	098552 parametecium
37	7	0.8	98	4	09UP45	09UP45 homo sapiens
38	7	0.8	105	17	096YV4	096YV4 sulfolobus
39	7	0.8	119	16	092CX1	092CX1 rickettsia
40	7	0.8	119	17	028742	028742 archaeoglobus
41	7	0.8	124	13	09PSQ5	09PSQ5 gallus galli
42	7	0.8	126	10	0947P8	0947P8 solanum tuberosum
43	7	0.8	129	16	098N52	098N52 rhizobium
44	7	0.8	129	16	08UE23	08UE23 agrobacterium
45	7	0.8	139	16	092G53	092G53 rhizobium
46	7	0.8	130	16	092DB6	092DB6 listeria monocytogenes
47	7	0.8	130	16	092DB6	092DB6 listeria monocytogenes
48	7	0.8	134	10	08Y8K1	08Y8K1 listeria monocytogenes
49	7	0.8	136	16	08W249	08W249 porphyra
50	7	0.8	136	16	045067	045067 bacillus subtilis
51	7	0.8	138	16	09K1N0	09K1N0 neisseria meningitidis
52	7	0.8	144	16	09JMW3	09JMW3 neisseria meningitidis
53	7	0.8	144	17	09YB59	09YB59 aeropyrum
54	7	0.8	156	10	09FCY2	09FCY2 arachidonic acid
55	7	0.8	157	10	09LNM0	09LNM0 halobacterium
56	7	0.8	157	17	09HJM0	09HJM0 halobacterium
57	7	0.8	161	4	09BS09	09BS09 homo sapiens
58	7	0.8	164	16	09X179	09X179 thermotoga
59	7	0.8	166	17	059317	059317 pyrococcus
60	7	0.8	167	17	08ZV66	08ZV66 pyrobaculum
61	7	0.8	169	17	08TVH6	08TVH6 methanopyrus
62	7	0.8	171	3	09HGM2	09HGM2 schizosaccharomyces
63	7	0.8	174	17	09HSP4	09HSP4 halobacterium
64	7	0.8	176	17	08U3R1	08U3R1 pyrococcus
65	7	0.8	177	10	09M9H3	09M9H3 arachidonic acid
66	7	0.8	177	16	0915D8	0915D8 pseudomonas
67	7	0.8	180	16	098G25	098G25 rhizobium
68	7	0.8	181	5	025405	025405 lymnaea stagnalis
69	7	0.8	186	16	08XX44	08XX44 ralsstonia solanaceae
70	7	0.8	191	10	0949H0	0949H0 hordeum vulgare
71	7	0.8	192	10	09XET4	09XET4 secale cereale
72	7	0.8	194	8	09MHV4	09MHV4 ascaris lumbricoides
73	7	0.8	194	8	09MHV3	09MHV3 baylisascaris
74	7	0.8	194	8	09MHV2	09MHV2 baylisascaris
75	7	0.8	194	8	09MHU3	09MHU3 parascaris
76	7	0.8	194	10	09M9G8	09M9G8 arachidonic acid
77	7	0.8	198	16	09F3H5	09F3H5 streptomyces
78	7	0.8	199	4	096NC0	096NC0 homo sapiens
79	7	0.8	199	11	09CPW7	09CPW7 mus musculus
80	7	0.8	200	13	09TAA6	09TAA6 gallus galli
81	7	0.8	202	16	09PBG7	09PBG7 xylella fastidiosa
82	7	0.8	204	2	086212	086212 enterococcus
83	7	0.8	206	10	09LVA5	09LVA5 arachidonic acid
84	7	0.8	206	16	097JH2	097JH2 clostridium
85	7	0.8	211	16	08XG83	08XG83 salmoneella
86	7	0.8	212	16	08RVU8	08RVU8 deinococcus
87	7	0.8	213	16	097T44	097T44 streptococcus
88	7	0.8	214	5	062188	062188 caenorhabditis
89	7	0.8	216	17	08T2J9	08T2J9 pyrococcus

90	7	0.8	217	16	092M70	092M70 rhizobium m	163	7	0.8	352	16	09KRT9	09KRT9 vibrio chol
91	7	0.8	224	16	09KTL1	09KTL1 neisseria m	164	7	0.8	355	10	09LE25	09LE25 chlamydom
92	7	0.8	225	16	097RT1	097RT1 streptococ	165	7	0.8	355	17	09HSA3	09HSA3 halobacteri
93	7	0.8	229	17	08ZRT9	08ZRT9 pyrobaculum	166	7	0.8	356	10	064647	064647 arabadopsis
94	7	0.8	230	2	0936W0	0936W0 pseudomonas	167	7	0.8	360	16	09KJ65	09KJ65 streptomyc
95	7	0.8	231	16	09HUC7	09HUC7 neisseria m	168	7	0.8	363	2	09EV21	09EV21 bradyrhizob
96	7	0.8	239	16	09JWY3	09JWY3 pseudomonas	169	7	0.8	365	13	09OZ18	09OZ18 cynops pyr
97	7	0.8	240	10	09JNO2	09JNO2 arabadopsis	170	7	0.8	368	16	09ZDE58	09ZDE58 listeria in
98	7	0.8	241	16	08YOD6	08YOD6 ralstonia s	171	7	0.8	368	16	08Y9D6	08Y9D6 listeria mo
99	7	0.8	242	16	08X770	08X770 escherichia	172	7	0.8	368	16	09KRT10	09KRT10 streptomyc
100	7	0.8	243	16	08ZPL3	08ZPL3 salmonella	173	7	0.8	370	5	001383	001383 drosophila
101	7	0.8	245	16	08Z6S1	08Z6S1 yersinia pe	174	7	0.8	371	16	09KE08	09KE08 bacillus ha
102	7	0.8	243	16	08Z6S1	08Z6S1 yersinia pe	175	7	0.8	371	16	098N04	098N04 rhizobium l
103	7	0.8	249	17	08ZRT5	08ZRT5 streptomyc	176	7	0.8	371	16	098ZG3	098ZG3 rhizobium l
104	7	0.8	251	2	09KJTK	09KJTK methanosa	177	7	0.8	371	16	08XNM2	08XNM2 clostridium
105	7	0.8	251	2	09KJTK	09KJTK methanosa	178	7	0.8	372	16	08XNM2	08XNM2 clostridium
106	7	0.8	251	2	09AMH3	09AMH3 bacillus me	179	7	0.8	379	16	09KRT31	09KRT31 streptomyc
107	7	0.8	251	2	09F5J0	09F5J0 streptomyc	180	7	0.8	379	16	09Z724	09Z724 chlamydia p
108	7	0.8	253	2	09F5J0	09F5J0 streptomyc	181	7	0.8	382	16	09CKD8	09CKD8 pasteurella
109	7	0.8	256	16	08ZS62	08ZS62 rhodobacter	182	7	0.8	383	2	056580	056580 vibrio chol
110	7	0.8	257	16	09KZB8	09KZB8 anabaena sp	183	7	0.8	383	2	056725	056725 vibrio para
111	7	0.8	260	2	093077	093077 streptomyc	184	7	0.8	385	16	09T3R2	09T3R2 pseudomonas
112	7	0.8	266	16	005853	005853 mycobacteri	185	7	0.8	389	5	09XV04	09XV04 caenorhabdi
113	7	0.8	274	6	08XNLO	08XNLO clostridium	186	7	0.8	389	16	09PGL7	09PGL7 xylella fas
114	7	0.8	274	6	09GME9	09GME9 archaeoglob	187	7	0.8	389	16	09CM70	09CM70 pasteurella
115	7	0.8	277	17	029076	029076 archaeoglob	188	7	0.8	392	2	09L8Q4	09L8Q4 pseudomonas
116	7	0.8	281	5	08SW14	08SW14 encephalito	189	7	0.8	392	5	08RTA3	08RTA3 drosophila
117	7	0.8	282	16	09ZTX1	09ZTX1 rhizobium m	190	7	0.8	394	5	08RTA3	08RTA3 drosophila
118	7	0.8	285	17	09ZTX0	09ZTX0 archaeoglob	191	7	0.8	395	4	09BSW2	09BSW2 trypanosoma
119	7	0.8	289	10	09ZTX0	09ZTX0 plasm sativ	192	7	0.8	395	16	P72972	P72972 synchocyst
120	7	0.8	290	11	09CTD9	09CTD9 mus musculi	193	7	0.8	396	16	034597	034597 bacillus su
121	7	0.8	291	5	09N5X2	09N5X2 caenorhabdi	194	7	0.8	397	4	043442	043442 homo sapien
122	7	0.8	293	16	083872	083872 treponema p	195	7	0.8	398	5	024466	024466 drosophila
123	7	0.8	295	5	09VTL8	09VTL8 drosophila	196	7	0.8	398	5	09Y000	09Y000 drosophila
124	7	0.8	296	16	09ZMM7	09ZMM7 rhizobium m	197	7	0.8	402	5	045792	045792 caenorhabdi
125	7	0.8	298	10	094B44	094B44 arabadopsis	198	7	0.8	402	16	08Z9J1	08Z9J1 salmonella
126	7	0.8	298	16	09KIU2	09KIU2 chlamydia p	199	7	0.8	402	16	08XCI2	08XCI2 escherichia
127	7	0.8	304	16	09ZSG3	09ZSG3 rhizobium m	200	7	0.8	402	16	08XBN3	08XBN3 escherichia
128	7	0.8	305	16	08UTB8	08UTB8 agrobacteri	201	7	0.8	404	16	09AK93	09AK93 streptomyc
129	7	0.8	306	16	09CT61	09CT61 pasteurella	202	7	0.8	408	5	08SUV7	08SUV7 encephalito
130	7	0.8	310	16	09CH39	09CH39 lactococcus	203	7	0.8	408	13	09PTK6	09PTK6 xenopus lae
131	7	0.8	313	5	08T895	08T895 clona intes	204	7	0.8	412	2	050371	050371 methylophil
132	7	0.8	313	16	09ZAX6	09ZAX6 listeria in	205	7	0.8	421	5	095086	095086 drosophila
133	7	0.8	314	16	08R6F5	08R6F5 fusobacteri	206	7	0.8	421	16	09EX44	09EX44 streptomyc
134	7	0.8	316	16	09EMP9	09EMP9 streptomyc	207	7	0.8	424	4	096SG3	096SG3 homo sapien
135	7	0.8	317	2	09AOK0	09AOK0 acinetobact	208	7	0.8	427	16	08R5N6	08R5N6 fusobacteri
136	7	0.8	317	16	09S5O7	09S5O7 rhizobium l	209	7	0.8	429	17	097A16	097A16 thermoplasm
137	7	0.8	317	16	08ZG9	08ZG9 pyrobaculum	210	7	0.8	430	10	09FT53	09FT53 arabadopsis
138	7	0.8	319	5	09NKJ7	09NKJ7 leishmania	211	7	0.8	432	16	08ZRT9	08ZRT9 salmonella
139	7	0.8	319	16	09BDY0	09BDY0 rhizobium l	212	7	0.8	433	5	091725	091725 salmonella
140	7	0.8	321	5	09VUD8	09VUD8 drosophila	213	7	0.8	433	11	054844	054844 mus musculi
141	7	0.8	322	16	08XIN3	08XIN3 clostridium	214	7	0.8	435	16	08XCM2	08XCM2 escherichia
142	7	0.8	323	16	08YU50	08YU50 anabaena sp	215	7	0.8	435	16	08XCM2	08XCM2 escherichia
143	7	0.8	323	16	08XFR7	08XFR7 salmonella	216	7	0.8	441	10	093YR3	093YR3 arabadopsis
144	7	0.8	323	17	08X9K0	08X9K0 escherichia	217	7	0.8	441	16	09ZYS9	09ZYS9 rhizobium m
145	7	0.8	323	17	09HKY3	09HKY3 thermoplasm	218	7	0.8	442	5	061670	061670 drosophila
146	7	0.8	325	10	09EYEL	09EYEL arabadopsis	219	7	0.8	442	10	09FL25	09FL25 arabadopsis
147	7	0.8	330	3	013303	013303 cryptococcu	220	7	0.8	443	10	09SC7	09SC7 ralstonia s
148	7	0.8	334	16	08XTN8	08XTN8 ralstonia s	221	7	0.8	446	16	08XTE2	08XTE2 homo sapien
149	7	0.8	337	16	08YXN0	08YXN0 anabaena sp	222	7	0.8	446	16	09C8E8	09C8E8 schizosacch
150	7	0.8	337	16	08XRS2	08XRS2 ralstonia s	223	7	0.8	449	4	043934	043934 homo sapien
151	7	0.8	338	3	012004	012004 saccharomye	224	7	0.8	449	4	09NX15	09NX15 homo sapien
152	7	0.8	339	5	09VJG9	09VJG9 drosophila	225	7	0.8	450	3	09P6K7	09P6K7 schizosacch
153	7	0.8	341	16	093MW7	093MW7 staphylococ	226	7	0.8	451	10	09C8E8	09C8E8 arabadopsis
154	7	0.8	343	4	08WU91	08WU91 homo sapien	227	7	0.8	452	16	08R6P4	08R6P4 fusobacteri
155	7	0.8	343	5	018014	018014 caenorhabdi	228	7	0.8	453	16	08WJUR	08WJUR brachionidi
156	7	0.8	343	16	09H2T3	09H2T3 pseudomonas	229	7	0.8	455	16	09XZL6	09XZL6 streptomyc
157	7	0.8	344	16	08UBN3	08UBN3 agrobacteri	230	7	0.8	456	2	09APM5	09APM5 ralstonia s
158	7	0.8	345	10	094JL0	094JL0 oryza sativ	231	7	0.8	458	4	09H6G4	09H6G4 bilophila v
159	7	0.8	347	16	08XBE9	08XBE9 escherichia	232	7	0.8	458	4	096RY4	096RY4 homo sapien
160	7	0.8	348	9	080301	080301 bacterioph	233	7	0.8	465	16	086374	086374 mycobacteri
161	7	0.8	351	2	030390	030390 neisseria m	234	7	0.8	471	16	09AA70	09AA70 caulobacter
162	7	0.8	351	10	09SYX6	09SYX6 nioclitana t	235	7	0.8	471	16	09AA70	09AA70 caulobacter

236	7	0.8	473	10	Q9M2X5	Q9M2X5 arabidopsis	309	7	0.8	643	17	Q97CP6	Q97CP6 thermoplasma
237	7	0.8	477	16	Q9JTN7	Q9Jtn7 neisseria m	310	7	0.8	648	16	Q8XYP3	Q8XYP3 ralstonia s
238	7	0.8	479	2	Q9S3Z3	Q9S3Z3 vibrio fisc	311	7	0.8	652	16	Q9KRP3	Q9KfB3 bacillus ha
239	7	0.8	479	16	Q9ZRS3	Q9Zrs3 rhizobium m	312	7	0.8	653	5	Q9VZK2	Q9vzr2 drosophila
240	7	0.8	480	2	054034	054034 pseudomonas	313	7	0.8	657	4	Q8TF57	Q8tf57 homo sapien
241	7	0.8	480	16	Q9HW85	Q9Hw85 pseudomonas	314	7	0.8	658	2	Q9ZHP9	Q9zhp9 streptomyc
242	7	0.8	485	16	Q9LOA2	Q9Loa2 streptomyc	315	7	0.8	664	5	Q24140	Q24140 drosophila
243	7	0.8	487	16	Q8X2N0	Q8xzn0 ralstonia s	316	7	0.8	673	4	Q96SH4	Q96sh4 homo sapien
244	7	0.8	491	2	Q93H06	Q93h06 streptomyc	317	7	0.8	676	2	Q9A158	Q9a158 burkholderi
245	7	0.8	493	10	Q8W0R0	Q8w0r0 sorghum bic	318	7	0.8	679	17	Q29619	Q29619 archaeoglob
246	7	0.8	494	17	Q96VC8	Q96vc8 sulfolobus	319	7	0.8	682	16	Q8YHK1	Q8yhk1 bruceella me
247	7	0.8	496	3	Q9URW9	Q9urw9 schizosacch	320	7	0.8	686	10	Q8S562	Q8s562 phaseolus v
248	7	0.8	496	11	Q9Z3M5	Q9z2m5 mus musculu	321	7	0.8	687	11	Q91XL2	Q91xl2 mus musculu
249	7	0.8	497	16	Q98BV6	Q98bv6 rhizobium l	322	7	0.8	693	3	Q94590	Q94590 schizosacch
250	7	0.8	497	16	Q98H29	Q98hz9 fusobacteri	323	7	0.8	695	5	Q9NJB3	Q9njb3 drosophila
251	7	0.8	499	16	Q8LYL9	Q8lyl9 anabaena sp	324	7	0.8	695	5	Q9NJB2	Q9njb2 drosophila
252	7	0.8	505	16	Q8UAA5	Q8uaa5 agrobacteri	325	7	0.8	695	5	Q9NJB1	Q9njb1 drosophila
253	7	0.8	512	2	052135	052135 escherichia	326	7	0.8	695	5	Q9NCB3	Q9ncb3 drosophila
254	7	0.8	512	2	Q93FL4	Q93fl4 citrobacter	327	7	0.8	695	5	Q9NCB2	Q9ncb2 drosophila
255	7	0.8	512	2	Q9AJ21	Q9aj21 escherichia	328	7	0.8	695	5	Q9NCB1	Q9ncb1 drosophila
256	7	0.8	512	2	047631	047631 escherichia	329	7	0.8	695	5	Q9NCB0	Q9ncb0 drosophila
257	7	0.8	512	16	Q85636	Q85636 escherichia	330	7	0.8	695	5	Q9NCA9	Q9nc9 drosophila
258	7	0.8	513	2	Q93HP2	Q93hp2 streptomyc	331	7	0.8	695	5	Q9NCA6	Q9nc6 drosophila
259	7	0.8	514	16	Q9ABT2	Q9abt2 caulobacter	332	7	0.8	695	5	Q9NCA5	Q9nc5 drosophila
260	7	0.8	516	10	Q9LN49	Q9ln49 arabidopsis	333	7	0.8	695	5	Q9NCA1	Q9nc1 drosophila
261	7	0.8	517	8	Q68912	Q68912 streptomyc	334	7	0.8	695	5	Q9NCA0	Q9nc0 drosophila
262	7	0.8	517	16	Q9ABX2	Q9abx2 rhizobium l	335	7	0.8	695	5	Q9N2R9	Q9n2r9 drosophila
263	7	0.8	518	16	Q8YCS7	Q8yxs7 bruceella me	336	7	0.8	695	5	Q9N2R8	Q9n2r8 drosophila
264	7	0.8	518	16	Q8YCS7	Q8yxs7 bruceella me	337	7	0.8	695	5	Q9N2R7	Q9n2r7 drosophila
265	7	0.8	519	12	Q8XWH5	Q8xwh5 ralstonia s	338	7	0.8	695	5	Q9N2R6	Q9n2r6 drosophila
266	7	0.8	523	16	Q33279	Q33279 rat cytoleg	339	7	0.8	695	5	Q9N2R5	Q9n2r5 drosophila
267	7	0.8	524	5	Q9XZ28	Q9xz28 mycobacteri	340	7	0.8	695	5	Q9N2R4	Q9n2r4 drosophila
268	7	0.8	528	2	Q9Z180	Q9z180 streptomyc	341	7	0.8	695	5	Q95WB5	Q95wb5 drosophila
269	7	0.8	528	16	Q9Z180	Q9z180 streptomyc	342	7	0.8	695	5	Q9NCA2	Q9nc2 drosophila
270	7	0.8	528	16	Q9Z180	Q9z180 streptomyc	343	7	0.8	695	5	Q9NCA3	Q9nc3 drosophila
271	7	0.8	533	17	Q95930	Q95930 pyrococcus	344	7	0.8	695	5	Q9NCA4	Q9nc4 drosophila
272	7	0.8	534	10	Q9A0U6	Q9a0u6 oryza sativ	345	7	0.8	695	5	Q9NCA7	Q9nc7 drosophila
273	7	0.8	548	5	001384	001384 drosophila	346	7	0.8	695	5	Q9NCA8	Q9nc8 drosophila
274	7	0.8	548	5	Q9VCA2	Q9vca2 drosophila	347	7	0.8	695	5	Q96441	Q96441 drosophila
275	7	0.8	555	4	Q8WXX5	Q8wxx5 homo sapien	348	7	0.8	695	5	Q9U8E8	Q9u8e8 drosophila
276	7	0.8	557	16	Q55534	Q55534 synechocyst	349	7	0.8	695	5	Q9U8E7	Q9u8e7 drosophila
277	7	0.8	559	17	Q9Y8U6	Q9y8u6 aeropyrum p	350	7	0.8	695	5	Q9U8E6	Q9u8e6 drosophila
278	7	0.8	566	10	Q9SK00	Q9sk00 arabidopsis	351	7	0.8	695	5	Q9U8E5	Q9u8e5 drosophila
279	7	0.8	566	16	Q9X1E2	Q9x1e2 thermotoga	352	7	0.8	695	5	Q9U8E4	Q9u8e4 drosophila
280	7	0.8	567	4	Q9H9K7	Q9h9k7 homo sapien	353	7	0.8	695	5	Q9U8E3	Q9u8e3 drosophila
281	7	0.8	567	4	Q96S06	Q96s06 homo sapien	354	7	0.8	695	5	Q9U8E2	Q9u8e2 drosophila
282	7	0.8	567	4	Q95R48	Q95r48 drosophila	355	7	0.8	695	5	Q9U8E1	Q9u8e1 drosophila
283	7	0.8	567	5	Q9VCA3	Q9vca3 drosophila	356	7	0.8	695	5	Q9U8D9	Q9u8d9 drosophila
284	7	0.8	567	16	Q912D3	Q912d3 pseudomonas	357	7	0.8	695	16	Q9ZWM4	Q9zwm4 rhizobium m
285	7	0.8	567	16	Q8ZJQ1	Q8zjq1 yersinia pe	358	7	0.8	698	17	Q26996	Q26996 methanobact
286	7	0.8	574	16	Q9A455	Q9a455 caulobacter	359	7	0.8	704	3	Q9P8D9	Q9p8d9 candida tro
287	7	0.8	581	17	Q96XHO	Q96xho sulfolobus	360	7	0.8	706	5	Q9VAL7	Q9val7 drosophila
288	7	0.8	582	12	Q9OH60	Q9oh60 gallia hepp	361	7	0.8	706	17	Q9YAG6	Q9yag6 aeropyrum p
289	7	0.8	584	5	061113	061113 trypanosoma	362	7	0.8	709	2	Q93CX3	Q93cx3 lactobacill
290	7	0.8	584	17	Q97Y06	Q97yq6 sulfolobus	363	7	0.8	712	16	Q8Y2S0	Q8y2s0 ralstonia s
291	7	0.8	588	9	Q64304	Q64304 bacterioph	364	7	0.8	717	4	Q9N052	Q9n052 homo sapien
292	7	0.8	593	16	Q9EY32	Q9ey32 xanthomonas	365	7	0.8	719	2	Q9P150	Q9p150 aquilex pyr
293	7	0.8	597	16	Q8U7M3	Q8u7m3 agrobacteri	366	7	0.8	719	5	Q9U4G1	Q9u4g1 drosophila
294	7	0.8	613	16	Q9PF38	Q9pf38 xyella fas	367	7	0.8	723	10	Q82670	Q82670 cicer ariet
295	7	0.8	613	16	Q9KVM0	Q9kvw0 vibrio chol	368	7	0.8	728	4	Q96MJ7	Q96mj7 homo sapien
296	7	0.8	614	16	Q97S07	Q97s07 streptococc	369	7	0.8	728	4	Q8TBM1	Q8tbl1 homo sapien
297	7	0.8	617	16	Q9PD27	Q9pdz7 xyella fas	370	7	0.8	728	11	Q70263	Q70263 mus musculu
298	7	0.8	623	16	Q8YEI3	Q8yei3 bruceella me	371	7	0.8	735	5	Q85XC4	Q85xc4 drosophila
299	7	0.8	627	10	Q49648	Q49648 arabidopsis	372	7	0.8	741	16	Q8UFT1	Q8uft1 agrobacteri
300	7	0.8	629	3	Q9P6U6	Q9p6u6 neosporea	373	7	0.8	744	13	Q91149	Q91149 notoptilham
301	7	0.8	629	5	Q16801	Q16801 paracentrot	374	7	0.8	745	12	Q36374	Q36374 alcelaphine
302	7	0.8	629	5	Q22720	Q22720 caenorhabdi	375	7	0.8	752	16	Q9PFZ5	Q9pfz5 xyella fas
303	7	0.8	630	10	Q9CAN0	Q9can0 arabidopsis	376	7	0.8	762	9	Q9ZKK6	Q9zkk6 bacterioph
304	7	0.8	630	11	Q99JV6	Q99jv6 mus musculu	377	7	0.8	765	16	Q8XN12	Q8xn12 ciostriidum
305	7	0.8	632	4	Q9BY20	Q9by20 homo sapien	378	7	0.8	766	2	Q9AER9	Q9aer9 pasteurella
306	7	0.8	639	4	Q9Y5K6	Q9y5k6 homo sapien	379	7	0.8	767	10	Q9SVF1	Q9svf1 arabidopsis
307	7	0.8	640	3	Q59863	Q59863 kluyveromyc	380	7	0.8	769	3	Q8X1E4	Q8x1e4 neurospora
308	7	0.8	641	2	Q9XC8	Q9xc8 streptomyc	381	7	0.8	781	16	Q92KG5	Q92kg5 rhizobium m

382	7	0.8	793	2	085356	085356 butyri/bri	455	7	0.8	1322	10	081490	081490 arabidopsis
383	7	0.8	793	10	092W18	092W18 chara coral	456	7	0.8	1326	5	09W4N4	09W4N4 drosophila
384	7	0.8	795	16	09KZN8	09KZN8 streptomyc	457	7	0.8	1330	6	097961	097961 vulpes vulp
385	7	0.8	802	5	09VJ74	09VJ74 drosophila	458	7	0.8	1335	5	09VE29	09VE29 drosophila
386	7	0.8	802	10	09ZTJ8	09ZTJ8 lycopersico	459	7	0.8	1336	4	014707	014707 homo sapien
387	7	0.8	804	16	09RXW4	09RXW4 delnoccoco	460	7	0.8	1373	4	09Y2F5	09Y2F5 homo sapien
388	7	0.8	811	9	09AZ46	09AZ46 coliphage K	461	7	0.8	1388	4	09Y4M4	09Y4M4 homo sapien
389	7	0.8	812	16	08XSQ1	08XSQ1 ralslonia s	462	7	0.8	1416	5	09VW1	09VW1 drosophila
390	7	0.8	815	10	064481	064481 arbidopsis	463	7	0.8	1475	10	09FZJ2	09FZJ2 arbidopsis
391	7	0.8	815	13	P87376	P87376 triturus ca	464	7	0.8	1526	5	019611	019611 caenorhabdi
392	7	0.8	820	4	043188	043188 homo sapien	465	7	0.8	1622	5	006530	006530 cryptospori
393	7	0.8	820	4	09BUQ8	09BUQ8 homo sapien	466	7	0.8	1643	5	09VJW1	09VJW1 drosophila
394	7	0.8	820	10	09LNL5	088661 rattus norv	467	7	0.8	1653	5	09NKE8	09NKE8 drosophila
395	7	0.8	822	11	088961	09btr3 homo sapien	468	7	0.8	1702	5	0950K2	0950K2 caenorhabdi
396	7	0.8	825	4	09HTR3	09HTR3 pseudomonas	469	7	0.8	1808	2	09EMA2	09EMA2 streptomyc
397	7	0.8	830	16	09HZY6	09HZY6 arbidopsis	470	7	0.8	1886	11	090Y81	090Y81 mus musculu
398	7	0.8	834	10	09FHU6	09H5V8 homo sapien	471	7	0.8	1914	13	091008	091008 gallus gall
399	7	0.8	836	4	09H5V8	096QY7 homo sapien	472	7	0.8	1939	16	08ZHV5	08ZHV5 yersinia pe
400	7	0.8	839	10	09LFF3	0911F3 arbidopsis	473	7	0.8	2024	2	09EMA3	09EMA3 streptomyc
401	7	0.8	846	4	075130	075130 homo sapien	474	7	0.8	2157	4	0960C6	0960C6 homo sapien
402	7	0.8	846	16	08ZE91	08ZE91 yersinia pe	475	7	0.8	2157	4	0960C6	0960C6 homo sapien
403	7	0.8	849	15	09YKR7	09YKR7 human immun	476	7	0.8	2197	12	088876	088876 tomato ring
404	7	0.8	850	15	08UTG3	08UTG3 human immun	477	7	0.8	2233	11	08S089	08S089 ratius norv
405	7	0.8	850	16	08YEL8	08YEL8 bruceella me	478	7	0.8	2653	12	09E1Z4	09E1Z4 cercopithec
406	7	0.8	851	15	0994H5	0994H5 human immun	479	7	0.8	2698	10	09X1R5	09X1R5 arbidopsis
407	7	0.8	852	15	070010	070010 human immun	480	7	0.8	3157	16	050378	050378 mycobacteri
408	7	0.8	857	15	0994R5	0994R5 human immun	481	7	0.8	3317	2	09RAH2	09RAH2 nostoc sp.
409	7	0.8	861	5	09NKK3	09NKK3 bomblyx mori	482	7	0.8	3716	16	08XS40	08XS40 mycobacteri
410	7	0.8	862	10	080468	080468 arbidopsis	483	7	0.8	6889	16	08XS40	08XS40 ralslonia s
411	7	0.8	862	10	09ZV48	09ZV48 arbidopsis	484	7	0.8	7829	5	018559	018559 caenorhabdi
412	7	0.8	862	16	09CBX9	09CBX9 mycobacteri	485	7	0.8	9507	2	09EWAI	09EWAI streptomyc
413	7	0.8	870	10	09RCDS	09RCDS arbidopsis	486	7	0.8	9510	2	093NX9	093NX9 streptomyc
414	7	0.8	872	2	093UQ2	093UQ2 alcaivoxax	487	6	0.7	10917	2	093NM6	093NM6 streptomyc
415	7	0.8	873	16	09PUC2	09PUC2 chlamydia m	488	6	0.7	20	15	085636	085636 moloney mur
416	7	0.8	887	16	08XY49	08XY49 ralslonia s	489	6	0.7	21	11	061917	061917 mus musculu
417	7	0.8	912	11	091XX6	091XX6 mus musculu	490	6	0.7	25	15	071927	071927 human immun
418	7	0.8	928	10	09HEU3	09HEU3 neurospora	491	6	0.7	31	4	090HM9	090HM9 homo sapien
419	7	0.8	947	10	023136	023136 arbidopsis	492	6	0.7	32	2	032493	032493 bacteroides
420	7	0.8	950	11	09WU36	09WU36 rattus norv	493	6	0.7	33	8	09T2M9	09T2M9 sus scrofa
421	7	0.8	961	9	08SBI1	08SBI1 bacterioph	494	6	0.7	33	13	09P872	09P872 rana catesb
422	7	0.8	983	3	09VE8	09VE8 ustilago ma	495	6	0.7	39	3	096T51	096T51 schizophyll
423	7	0.8	985	12	09ORA6	09ORA6 tomato ring	496	6	0.7	42	2	048078	048078 mycoplasma
424	7	0.8	985	12	09ORA7	09ORA7 tomato ring	497	6	0.7	43	4	090J35	090J35 homo sapien
425	7	0.8	985	12	09ORA8	09ORA8 tomato ring	498	6	0.7	44	6	09RTN7	09RTN7 sus scrofa
426	7	0.8	985	12	09ORA8	09ORA8 tomato ring	499	6	0.7	45	11	091ZQ3	091ZQ3 cavia porce
427	7	0.8	985	12	09ORA8	09ORA8 tomato ring	500	6	0.7	46	2	048501	048501 lactobacill
428	7	0.8	987	3	09USU1	09USU1 schizosacch	501	6	0.7	47	2	08VLZ0	08VLZ0 treponema d
429	7	0.8	992	3	036025	036025 salmoneila	502	6	0.7	47	8	095CL2	095CL2 podolobium
430	7	0.8	993	16	08Z745	08Z745 rice yellow	503	6	0.7	47	8	095CL1	095CL1 podolobium
431	7	0.8	996	16	08U7Y7	08U7Y7 agrobacteri	504	6	0.7	47	8	095CK9	095CK9 podolobium
432	7	0.8	999	12	08E525	08E525 rice yellow	505	6	0.7	47	8	095CK6	095CK6 pulitanea d
433	7	0.8	999	12	08E519	08E519 escherichia	506	6	0.7	48	2	09KJ74	09KJ74 scytonema h
434	7	0.8	1005	16	08X9L4	08X9L4 arbidopsis	507	6	0.7	48	4	095743	095743 homo sapien
435	7	0.8	1006	10	08RXU9	08RXU9 peanut stun	508	6	0.7	49	10	039226	039226 arbidopsis
436	7	0.8	1006	12	085064	085064 schistosoma	509	6	0.7	51	1	09HH93	09HH93 sulfolobus
437	7	0.8	1011	5	096527	096527 glycine max	510	6	0.7	52	4	09PIN5	09PIN5 homo sapien
438	7	0.8	1012	10	09LKR4	09LKR4 drosophila	511	6	0.7	52	16	09PGV4	09PGV4 xylella fas
439	7	0.8	1028	5	09VJC4	09VJC4 drosophila	512	6	0.7	54	2	092ND7	092ND7 thermus the
440	7	0.8	1031	5	09VRE9	09VRE9 drosophila	513	6	0.7	55	13	09PVQ6	09PVQ6 otyzias lat
441	7	0.8	1033	11	09WTW1	09WTW1 rattus norv	514	6	0.7	55	16	08YMA2	08YMA2 anabena sp
442	7	0.8	1050	10	09LGH9	09LGH9 oryza sativ	515	6	0.7	55	16	08YMA2	08YMA2 anabena sp
443	7	0.8	1127	5	08T519	08T519 anophelis g	516	6	0.7	56	2	09L7F8	09L7F8 synechococc
444	7	0.8	1132	5	002432	002432 drosophila	517	6	0.7	57	2	09ZAT1	09ZAT1 streptococc
445	7	0.8	1132	5	09VMA3	09VMA3 drosophila	518	6	0.7	58	2	09ZAT1	09ZAT1 streptococc
446	7	0.8	1138	16	092XB6	092XB6 rhizobium m	519	6	0.7	59	4	096074	096074 homo sapien
447	7	0.8	1157	16	097F62	097F62 clostridium	520	6	0.7	59	5	08WPZ4	08WPZ4 plasmodium
448	7	0.8	1187	13	012965	012965 fuigu rubrip	521	6	0.7	60	10	093387	093387 nicotiana t
449	7	0.8	1208	5	09U4W1	09U4W1 aedes aegy	522	6	0.7	61	2	054538	054538 streptococc
450	7	0.8	1223	3	002979	002979 saccharomyc	523	6	0.7	61	2	093PR1	093PR1 bradyrhizob
451	7	0.8	1236	5	09GRM1	09GRM1 leishmania	524	6	0.7	63	16	09W7X7	09W7X7 arbidopsis
452	7	0.8	1286	10	080725	080725 arbidopsis	525	6	0.7	64	16	09PD64	09PD64 xylella fas
453	7	0.8	1292	10	09M109	09M109 arbidopsis	526	6	0.7	64	16	09PD64	09PD64 xylella fas
454	7	0.8	1300	4	013999	013999 homo sapien	527	6	0.7	66	16	08X3N6	08X3N6 escherichia

528	6	0.7	67	16	08XSY4	08XSY4 ralstonia s	601	6	0.7	89	2	045189	045189 borrelia ga
529	6	0.7	68	2	09Z440	09Z440 pseudomonas	602	6	0.7	89	2	045190	045190 borrelia ga
530	6	0.7	68	6	028692	028692 oryctolagus	603	6	0.7	89	2	045188	045188 borrelia ga
531	6	0.7	68	16	082166	082166 yersinia pe	604	6	0.7	89	16	08REB6	08REB6 fusobacteri
532	6	0.7	69	2	0917G2	0917G2 synechococ	605	6	0.7	89	17	09HPU4	09HPU4 halobacteri
533	6	0.7	69	2	051902	051902 proteus mir	606	6	0.7	89	17	0972M7	0972M7 sulfolobus
534	6	0.7	69	5	09VN37	09VN37 drosophila	607	6	0.7	90	2	032473	032473 pseudomonas
535	6	0.7	70	5	09VRI6	09VRI6 drosophila	608	6	0.7	90	5	P91520	P91520 caenorhabdi
536	6	0.7	70	12	09E4B9	09E4B9 hepatitis b	609	6	0.7	90	11	08RA12	08RA12 mus musculi
537	6	0.7	70	12	067720	067720 human adeno	610	6	0.7	91	2	056426	056426 thermus the
538	6	0.7	71	2	09L710	09L710 synechococ	611	6	0.7	91	4	014075	014075 homo sapien
539	6	0.7	72	2	091RB3	091RB3 plesiomonas	612	6	0.7	91	5	09VC9	09VC9 drosophila
540	6	0.7	72	12	091SA0	091SA0 regina rana	613	6	0.7	91	10	048605	048605 hordem vul
541	6	0.7	72	15	083373	083373 murine leuk	614	6	0.7	91	15	056583	056583 mlv-like en
542	6	0.7	72	16	092JL4	092JL4 rickettsia	615	6	0.7	91	15	0926G7	0926G7 rickettsia
543	6	0.7	73	2	085925	085925 sphingomona	616	6	0.7	91	16	09RY58	09RY58 delnoccocus
544	6	0.7	73	13	073777	073777 gallus gall	617	6	0.7	92	10	08W217	08W217 setaria fab
545	6	0.7	73	16	08R5T5	08R5T5 thermomater	618	6	0.7	92	10	08W214	08W214 setaria ita
546	6	0.7	74	4	09NRX6	09NRX6 homo sapien	619	6	0.7	92	10	08W2H4	08W2H4 setaria ver
547	6	0.7	74	8	037408	037408 emericeia	620	6	0.7	92	16	09K6W5	09K6W5 bacillus ha
548	6	0.7	74	8	037404	037404 allomyces m	621	6	0.7	92	16	053745	053745 mycobacteri
549	6	0.7	74	10	08SAY4	08SAY4 oryza sativ	622	6	0.7	93	2	066021	066021 pseudomonas
550	6	0.7	74	11	09DB63	09DB63 mus musculi	623	6	0.7	93	8	035964	035964 oryza sativ
551	6	0.7	74	16	09AB64	09AB64 caulobacter	624	6	0.7	93	16	08XEA4	08XEA4 escherichia
552	6	0.7	75	16	08X448	08X448 escherichia	625	6	0.7	93	16	P76067	P76067 escherichia
553	6	0.7	75	2	09AFZ2	09AFZ2 shigella fl	626	6	0.7	94	2	09F2D0	09F2D0 salmonella
554	6	0.7	76	2	09JMM3	09JMM3 bradyrhizob	627	6	0.7	94	10	09FTL0	09FTL0 oryza sativ
555	6	0.7	76	10	09SH04	09SH04 arabidopsis	628	6	0.7	94	15	09DPY7	09DPY7 human immun
556	6	0.7	76	16	08XUW0	08XUW0 ralstonia s	629	6	0.7	94	15	09DPX6	09DPX6 human immun
557	6	0.7	76	17	0974F2	0974F2 sulfolobus	630	6	0.7	94	16	P73959	P73959 synechocyst
558	6	0.7	77	5	09VP08	09VP08 drosophila	631	6	0.7	94	17	08RP49	08RP49 methanocyst
559	6	0.7	77	17	08Z211	08Z211 pyrobaculum	632	6	0.7	95	2	007153	007153 mycobacteri
560	6	0.7	78	2	08VSH9	08VSH9 shigella fl	633	6	0.7	95	12	08VAT1	08VAT1 white spot
561	6	0.7	78	16	08VJN9	08VJN9 mycobacteri	634	6	0.7	95	15	003458	003458 feline leuk
562	6	0.7	78	17	028207	028207 archaeoglob	635	6	0.7	95	17	08ZUH7	08ZUH7 pyrobaculum
563	6	0.7	79	2	09FPD1	09FPD1 borrelia bu	636	6	0.7	96	2	093705	093705 salmonella
564	6	0.7	79	16	09CC09	09CC09 mycobacteri	637	6	0.7	96	16	080801	080801 agrobacteri
565	6	0.7	79	16	08ZJK0	08ZJK0 yersinia pe	638	6	0.7	96	17	08RYM5	08RYM5 methanosarc
566	6	0.7	79	16	08R968	08R968 thermomater	639	6	0.7	97	2	068308	068308 aeromonas s
567	6	0.7	80	12	041946	041946 murid herpe	640	6	0.7	98	10	08S4X8	08S4X8 oryza sativ
568	6	0.7	80	16	08Y3E4	08Y3E4 ralstonia s	641	6	0.7	98	16	09Z8X3	09Z8X3 chlamydia p
569	6	0.7	81	2	085044	085044 thiodaciliu	642	6	0.7	98	17	09HPX8	09HPX8 halobacteri
570	6	0.7	81	2	048496	048496 lactobacilli	643	6	0.7	99	10	094AN0	094AN0 arabidopsis
571	6	0.7	81	10	024195	024195 oryza sativ	644	6	0.7	100	5	08SRK5	08SRK5 encephalito
572	6	0.7	81	10	024205	024205 oryza sativ	645	6	0.7	100	15	P89820	P89820 human immun
573	6	0.7	81	15	P90284	P90284 human immun	646	6	0.7	100	15	P89820	P89820 human immun
574	6	0.7	81	15	P90285	P90285 human immun	647	6	0.7	100	17	09Y9S0	09Y9S0 human immun
575	6	0.7	81	16	098PM3	098PM3 rhizobium i	648	6	0.7	101	2	051826	051826 aeropyrum p
576	6	0.7	81	16	08ZGW6	08ZGW6 yersinia pe	649	6	0.7	101	5	09VP77	09VP77 shigella fl
577	6	0.7	82	16	091625	091625 pseudomonas	650	6	0.7	101	5	08SX73	08SX73 drosophila
578	6	0.7	82	16	08YFK4	08YFK4 bruceella me	651	6	0.7	101	12	08UYE7	08UYE7 tt virus. o
579	6	0.7	83	7	09G1J9	09G1J9 damuliscus	652	6	0.7	102	2	054771	054771 synechococ
580	6	0.7	83	16	09A931	09A931 caulobacter	653	6	0.7	102	11	064098	064098 mus sp. tru
581	6	0.7	84	17	08TRP6	08TRP6 methanosarc	654	6	0.7	102	16	09EAV3	09EAV3 rhizobium i
582	6	0.7	84	6	095LE7	095LE7 canis famli	655	6	0.7	103	2	09FPD29	09FPD29 burkholderi
583	6	0.7	85	4	09H513	09H513 homo sapien	656	6	0.7	103	10	004824	004824 tipuana tip
584	6	0.7	85	13	09PWL7	09PWL7 brachydanio	657	6	0.7	103	13	090224	090224 aplocheilic
585	6	0.7	85	16	08RGY2	08RGY2 fusobacteri	658	6	0.7	103	16	08YVK8	08YVK8 aplocheilic
586	6	0.7	86	16	091Y83	091Y83 neisseria m	659	6	0.7	104	2	09A0W8	09A0W8 shigella fl
587	6	0.7	86	16	08UHD5	08UHD5 lactococcus	660	6	0.7	104	10	09FSQ4	09FSQ4 oryza sativ
588	6	0.7	87	2	048505	048505 lactococcus	661	6	0.7	104	13	090242	090242 aplocheilic
589	6	0.7	87	2	093NB3	093NB3 lactococcus	662	6	0.7	104	15	091929	091929 human immun
590	6	0.7	87	3	001298	001298 neurospora	663	6	0.7	104	16	099T07	099T07 human immun
591	6	0.7	87	15	090AC1	090AC1 human immun	664	6	0.7	104	17	08ZU46	08ZU46 pyrobaculum
592	6	0.7	87	16	08ZLP6	08ZLP6 salmonella	665	6	0.7	105	2	033677	033677 staphylococ
593	6	0.7	87	16	08Z3D9	08Z3D9 salmonella	666	6	0.7	105	2	032926	032926 mycobacteri
594	6	0.7	87	17	08TUZ0	08TUZ0 methanopyru	667	6	0.7	105	6	007708	007708 staphylococ
595	6	0.7	88	6	046444	046444 dasyurus al	668	6	0.7	105	6	08WNX4	08WNX4 oryctolagus
596	6	0.7	88	10	08W216	08W216 setaria ita	669	6	0.7	105	16	099SM1	099SM1 staphylococ
597	6	0.7	88	16	09KGR2	09KGR2 bacillus ha	670	6	0.7	106	16	09A4D2	09A4D2 caulobacter
598	6	0.7	88	16	0929C0	0929C0 listeria in	671	6	0.7	106	16	08ZFD7	08ZFD7 yersinia pe
599	6	0.7	88	16	08Y522	08Y522 listeria mo	672	6	0.7	106	16	08RYO0	08RYO0 anabaena sp
600	6	0.7	88	16	08R763	08R763 thermomater	673	6	0.7	106	16	09X9V4	09X9V4 streptomyce

674	6	0.7	107	9	Q9FZW8	Q9fzw8 bacterioph	747	6	0.7	119	9	Q9B081	Q9b081 mycobacteri
675	6	0.7	107	16	P71950	P71950 mycobacteri	748	6	0.7	119	16	Q9HXX8	Q9hxx8 pseudomonas
676	6	0.7	107	16	Q9L0H4	Q9l0h4 streptomyc	749	6	0.7	119	16	Q8XP23	Q8xp23 clostridium
677	6	0.7	107	17	Q9Y753	Q9y753 sulfolobus	750	6	0.7	120	2	Q8RP27	Q8rp27 shigella fl
678	6	0.7	108	4	Q9Y533	Q9y533 homo sapien	751	6	0.7	120	5	Q9VJG6	Q9vjg6 drosophila
679	6	0.7	108	5	Q9VE99	Q9ve99 drosophila	752	6	0.7	120	5	Q9VJG6	Q9vjg6 drosophila
680	6	0.7	108	9	Q64354	Q64354 bacterioph	753	6	0.7	120	10	Q8S1U1	Q8s1u1 oryza sativ
681	6	0.7	108	10	Q9LXMO	Q9lxm0 arabidopsis	754	6	0.7	120	15	Q8R243	Q8r243 oryza sativ
682	6	0.7	108	16	Q9A316	Q9a316 caulobacter	755	6	0.7	120	15	Q90DY5	Q90dy5 porcine end
683	6	0.7	109	2	Q50470	Q50470 neisseria g	756	6	0.7	120	16	Q9HJZ0	Q9hjz0 pseudomonas
684	6	0.7	109	4	Q8WY77	Q8wy77 homo sapien	757	6	0.7	120	16	Q8ZKJ9	Q8zkj9 salmonella
685	6	0.7	109	5	Q61467	Q61467 caenorhabdi	758	6	0.7	120	16	Q8ZSM7	Q8zsm7 salmonella
686	6	0.7	109	8	Q9T4C7	Q9t4c7 neiphoselmi	759	6	0.7	120	17	Q8Z277	Q8z277 pyrobaculum
687	6	0.7	109	12	Q9WAY5	Q9way5 tt virus. o	760	6	0.7	121	8	Q9TLM7	Q9tlm7 cyanidium c
688	6	0.7	109	16	Q9XE22	Q9xe22 escherichia	761	6	0.7	121	16	Q9K662	Q9k662 bacillus ha
689	6	0.7	109	16	Q8X4W8	Q8x4w8 escherichia	762	6	0.7	122	8	Q9MPC4	Q9mpc4 delta vulgar
690	6	0.7	109	17	Q9YG06	Q9yg06 aeropyrum p	763	6	0.7	122	10	Q9LH88	Q9lh88 arabidopsis
691	6	0.7	110	1	P95922	P95922 sulfolobus	764	6	0.7	122	16	Q9HDP8	Q9hdp8 rhizobium l
692	6	0.7	110	5	Q94778	Q94778 trypanosoma	765	6	0.7	123	3	Q9HDP7	Q9hdp7 colistotric
693	6	0.7	110	6	Q46447	Q46447 cercarictus	766	6	0.7	123	16	Q8YC18	Q8yc18 rhizobium m
694	6	0.7	110	9	Q8W6M5	Q8w6m5 sinorhizobl	767	6	0.7	124	2	Q68464	Q68464 brucella me
695	6	0.7	110	10	Q9AKX7	Q9aix7 oryza sativ	768	6	0.7	124	5	Q9U027	Q9u027 giardia lam
696	6	0.7	110	16	Q8XAN0	Q8xan0 escherichia	769	6	0.7	124	5	Q9V333	Q9v333 drosophila
697	6	0.7	111	10	Q42251	Q42251 arabidopsis	770	6	0.7	124	16	Q97ON0	Q97on0 streptococc
698	6	0.7	111	12	Q990T2	Q990t2 hepatitis c	771	6	0.7	124	16	Q8ZNB7	Q8znb7 salmonella
699	6	0.7	111	17	Q9YF60	Q9yfe0 aeropyrum p	772	6	0.7	124	16	Q8Z4Z8	Q8z4z8 salmonella
700	6	0.7	112	3	Q9KX29	Q9kx29 pseudomonas	773	6	0.7	124	17	Q28447	Q28447 archaeoglob
701	6	0.7	112	3	Q05662	Q05662 saccharomyc	774	6	0.7	125	2	Q68467	Q68467 rickettsia
702	6	0.7	112	4	Q9BMU8	Q9bm8 homo sapien	775	6	0.7	125	16	Q8Z350	Q8z350 salmonella
703	6	0.7	112	4	Q9B216	Q9b216 homo sapien	776	6	0.7	125	17	Q97X82	Q97x82 sulfolobus
704	6	0.7	112	12	Q99355	Q99355 banana mild	777	6	0.7	126	2	Q9EPY3	Q9ep93 neisseria c
705	6	0.7	112	12	Q8UYB9	Q8uyb9 tt virus. o	778	6	0.7	126	5	Q9AMU6	Q9am6 bradyrhizob
706	6	0.7	113	16	Q8YRC7	Q8yrc7 anabaena sp	779	6	0.7	126	5	Q8SS64	Q8ss64 encephalito
707	6	0.7	113	4	Q9BFX3	Q9bfx3 homo sapien	780	6	0.7	126	16	Q9A8U8	Q9a8u8 caulobacter
708	6	0.7	113	4	Q9BMV0	Q9bmv0 homo sapien	781	6	0.7	126	16	Q9KRC9	Q9krc9 streptomyc
709	6	0.7	113	4	Q9BMU7	Q9bm9 homo sapien	782	6	0.7	126	16	Q8XT22	Q8xt22 raistonia s
710	6	0.7	113	10	Q81232	Q81232 zea mays (m	783	6	0.7	127	4	Q96L29	Q96l29 homo sapien
711	6	0.7	113	11	Q921U3	Q921u3 mus musculu	784	6	0.7	127	5	Q9NB70	Q9nb70 dicystostell
712	6	0.7	113	10	Q921U3	Q921u3 listeria in	785	6	0.7	127	5	Q8RTB8	Q8rtb8 mesocricetu
713	6	0.7	113	16	Q926A0	Q926a0 streptomyc	786	6	0.7	127	11	P97282	P97282 mycobacteri
714	6	0.7	113	16	Q92589	Q92589 salmonella	787	6	0.7	127	16	Q33192	Q33192 anopheles a
715	6	0.7	114	2	Q87647	Q87647 mycobacteri	788	6	0.7	128	3	Q9HFP7	Q9hfp7 coccidioid
716	6	0.7	114	2	Q9X5X5	Q9x5x5 mycobacteri	789	6	0.7	128	5	Q16918	Q16918 anopheles a
717	6	0.7	114	2	Q9R9D9	Q9r9d9 bacillus su	790	6	0.7	128	10	Q9VY96	Q9vy96 arabidopsis
718	6	0.7	114	3	Q08958	Q08958 leavenworth	791	6	0.7	128	12	Q8VA19	Q8va19 white spot
719	6	0.7	114	10	Q9AKR3	Q9akr3 leavenworth	792	6	0.7	128	15	Q85652	Q85652 murine leuk
720	6	0.7	114	10	Q9AKR5	Q9akr5 leavenworth	793	6	0.7	128	16	Q8ZID9	Q8zid9 yersinia pe
721	6	0.7	114	10	Q9AKR4	Q9akr4 leavenworth	794	6	0.7	129	10	Q9FRI9	Q9fri9 arabidopsis
722	6	0.7	114	10	Q9AXR3	Q9axr3 leavenworth	795	6	0.7	129	16	Q8YHNS	Q8yhn5 bruceella me
723	6	0.7	114	10	Q9AXR1	Q9axr1 leavenworth	796	6	0.7	129	16	Q52884	Q52884 rhizobium m
724	6	0.7	114	10	Q9SIC8	Q9sic8 arabidopsis	797	6	0.7	129	16	Q92K41	Q92k41 rhizobium m
725	6	0.7	114	10	Q8S5V7	Q8s5v7 oryza sativ	798	6	0.7	130	12	Q9JFMO	Q9jfm0 human adeno
726	6	0.7	114	15	Q36580	Q36580 multiple sc	799	6	0.7	130	12	Q9JFMS	Q9jfm3 human adeno
727	6	0.7	114	16	Q9AAS5	Q9aas5 caulobacter	800	6	0.7	130	17	Q29186	Q29186 archaeoglob
728	6	0.7	115	2	Q9JPE1	Q9jpe1 neisseria s	801	6	0.7	131	2	Q9S387	Q9s387 listeria mo
729	6	0.7	115	2	Q8VTC5	Q8vtc5 neisseria m	802	6	0.7	131	11	Q9DP63	Q9dp63 mus musculu
730	6	0.7	115	2	Q8RM12	Q8rm12 clostridium	803	6	0.7	131	16	Q9RTU3	Q9rty3 deinococcus
731	6	0.7	115	8	Q98S99	Q98s99 guillardia	804	6	0.7	131	16	Q9AVV1	Q9avv1 caulobacter
732	6	0.7	115	8	Q98S99	Q98s99 guillardia	805	6	0.7	131	16	Q8ZAV3	Q8zav3 yersinia pe
733	6	0.7	115	10	Q43762	Q43762 hordium vul	806	6	0.7	132	4	Q9UG94	Q9ug94 homo sapien
734	6	0.7	115	16	Q49691	Q49691 mycobacteri	807	6	0.7	132	4	Q9UG94	Q9ug94 homo sapien
735	6	0.7	116	2	Q93JP9	Q93jp9 ruminococcu	808	6	0.7	132	5	Q9NPA1	Q9npa1 plasmodium
736	6	0.7	116	10	Q9AS93	Q9as93 oryza sativ	809	6	0.7	132	5	Q9NPA1	Q9npa1 plasmodium
737	6	0.7	116	12	Q91F84	Q91f84 chilo iride	810	6	0.7	132	11	Q61638	Q61638 mus musculu
738	6	0.7	116	16	Q9HYO0	Q9hyo0 pseudomonas	811	6	0.7	133	9	Q9U241	Q9u241 caenorhabdi
739	6	0.7	117	12	Q8VYD2	Q8vyd2 tt virus. o	812	6	0.7	133	11	Q9C3M0	Q9c3m0 straphylococ
740	6	0.7	117	12	Q8UYC8	Q8uyc8 tt virus. o	813	6	0.7	133	12	Q9CYO8	Q9cyo8 mus musculu
741	6	0.7	118	2	Q9APX6	Q9apx6 pseudomonas	814	6	0.7	133	16	Q8YMX4	Q8ymx4 anabaena sp
742	6	0.7	118	2	Q9FCY9	Q9fcy9 erwinia ste	815	6	0.7	133	16	Q8YMX4	Q8ymx4 anabaena sp
743	6	0.7	118	2	Q46672	Q46672 escherichia	816	6	0.7	133	16	Q8YMX4	Q8ymx4 anabaena sp
744	6	0.7	118	6	Q95JV4	Q95jv4 macaca fasc	817	6	0.7	133	16	Q93JB4	Q93jb4 streptomyc
745	6	0.7	118	6	Q95JV4	Q95jv4 macaca fasc	818	6	0.7	133	16	Q93JB4	Q93jb4 streptomyc
746	6	0.7	118	17	Q8Z248	Q8z248 pyrobaculum	819	6	0.7	133	16	Q98QN4	Q98qn4 mycoplasma

820	6	0.7	133	16	092P06	092P06 rhizobium m	893	6	0.7	143	12	091N36	091N36 hepatitis b
821	6	0.7	134	2	09K2H4	09K2H4 bacillus li	894	6	0.7	143	16	093R87	093R87 salmonella
822	6	0.7	134	2	09K2H3	09K2H3 bacillus so	895	6	0.7	143	16	08XAT8	08XAT8 escherichia
823	6	0.7	134	2	09L788	09L788 bacillus so	896	6	0.7	143	17	09TFH9	09TFH9 aeropyrum p
824	6	0.7	134	2	09L787	09L787 bacillus so	897	6	0.7	144	10	09CAM0	09CAM0 arabisdopsis
825	6	0.7	134	2	09L786	09L786 bacillus li	898	6	0.7	145	6	09BF81	09BF81 didelphis m
826	6	0.7	134	2	049963	049963 mycobacteri	899	6	0.7	145	6	09BF80	09BF80 macropus eu
827	6	0.7	134	3	0960R4	0960R4 coccidioidi	900	6	0.7	145	10	08S181	08S181 oryza sativ
828	6	0.7	134	4	09NV29	09NV29 homo sapien	901	6	0.7	145	17	027382	027382 methanobact
829	6	0.7	134	4	096FZ0	096FZ0 homo sapien	902	6	0.7	146	2	09EVZ8	09EVZ8 streptococc
830	6	0.7	134	9	003934	003934 bacterioph	903	6	0.7	146	5	09BP06	09BP06 metagonimus
831	6	0.7	134	16	08UGT2	08UGT2 agrobacteri	904	6	0.7	146	9	09B0B9	09B0B9 mycobacteri
832	6	0.7	134	17	08ZMQ0	08ZMQ0 pyrobaculum	905	6	0.7	146	10	09LPA9	09LPA9 arabisdopsis
833	6	0.7	135	16	09X703	09X703 corynebacte	906	6	0.7	146	16	09JXV2	09JXV2 neisseria m
834	6	0.7	135	16	09A908	09A908 caulobacte	907	6	0.7	146	16	09JXV1	09JXV1 neisseria m
835	6	0.7	135	16	09XA32	09XA32 streptomyce	908	6	0.7	146	16	09H235	09H235 pseudomonas
836	6	0.7	135	17	08U0E7	08U0E7 pyrococcus	909	6	0.7	146	16	08H046	08H046 streptomyce
837	6	0.7	136	5	09NMF9	09NMF9 leishmania	910	6	0.7	147	2	08S781	08S781 myxococcus
838	6	0.7	136	5	041859	041859 zea mays (m	911	6	0.7	147	4	09H657	09H657 homo sapien
839	6	0.7	136	10	09D8S0	09D8S0 mus muscul	912	6	0.7	147	11	09CRM2	09CRM2 mus muscul
840	6	0.7	136	11	09D8S0	09D8S0 mus muscul	913	6	0.7	147	16	08UEC6	08UEC6 agrobacteri
841	6	0.7	136	12	08QVW3	08QVW3 sen virus.	914	6	0.7	148	2	09S3M2	09S3M2 bordetella
842	6	0.7	136	12	08QVW2	08QVW2 sen virus.	915	6	0.7	148	2	045805	045805 clostridium
843	6	0.7	136	12	08QVW1	08QVW1 sen virus.	916	6	0.7	148	2	052867	052867 clostridium
844	6	0.7	137	2	054029	054029 propionigen	917	6	0.7	148	9	038345	038345 lactococcus
845	6	0.7	137	2	086989	086989 proteus mir	918	6	0.7	148	10	09SKH0	09SKH0 arabisdopsis
846	6	0.7	137	16	09R2T7	09R2T7 delinococcus	919	6	0.7	148	10	09FTL3	09FTL3 arabisdopsis
847	6	0.7	138	3	09A1J3	09A1J3 escherichia	920	6	0.7	148	12	091BS1	091BS1 turkey heip
848	6	0.7	138	10	0932C6	0932C6 ashuya goss	921	6	0.7	148	16	09CLS6	09CLS6 pasteurella
849	6	0.7	138	12	08ON95	08ON95 arabisdopsis	922	6	0.7	148	16	0970S3	0970S3 streptococc
850	6	0.7	138	13	093606	093606 ectocarpus	923	6	0.7	148	17	097Y63	097Y63 sulfolobus
851	6	0.7	138	16	08X603	08X603 escherichia	924	6	0.7	149	5	096601	096601 caenorhabdi
852	6	0.7	138	17	097WZ9	097WZ9 sulfolobus	925	6	0.7	149	10	09M209	09M209 arabisdopsis
853	6	0.7	139	8	094RU8	094RU8 oncorhynch	926	6	0.7	149	16	08YR04	08YR04 anabena sp
854	6	0.7	139	8	094RU6	094RU6 microplitis	927	6	0.7	149	16	09KXK8	09KXK8 streptomyce
855	6	0.7	139	9	064352	064352 bacterioph	928	6	0.7	150	2	093PM5	093PM5 streptococc
856	6	0.7	139	9	092X64	092X64 mycobacteri	929	6	0.7	150	16	025203	025203 helicobacte
857	6	0.7	139	9	006913	006913 brassica na	930	6	0.7	150	16	09PM45	09PM45 campylobact
858	6	0.7	139	10	08VZK3	08VZK3 chlamydomon	931	6	0.7	150	16	09A7J5	09A7J5 caulobacter
859	6	0.7	139	10	08VZK3	08VZK3 chlamydomon	932	6	0.7	151	2	047630	047630 escherichia
860	6	0.7	139	16	09A8A7	09A8A7 caulobacter	933	6	0.7	151	2	091A89	091A89 aeromonas h
861	6	0.7	139	16	092JY2	092JY2 rhizobium m	934	6	0.7	151	2	093FL3	093FL3 citrobacter
862	6	0.7	139	17	08TKU2	08TKU2 methanosarc	935	6	0.7	151	4	09H6C7	09H6C7 homo sapien
863	6	0.7	140	2	09FTW3	09FTW3 streptomyce	936	6	0.7	151	5	08SYR4	08SYR4 drosophila
864	6	0.7	140	2	09EUL8	09EUL8 corynebacte	937	6	0.7	152	16	09KRU5	09KRU5 vibrio chol
865	6	0.7	140	2	09FA65	09FA65 azoarcus ev	938	6	0.7	152	16	091S25	091S25 pseudomonas
866	6	0.7	140	2	09EX95	09EX95 corynebacte	939	6	0.7	153	2	09F7S1	09F7S1 uncultured
867	6	0.7	140	2	08VOY9	08VOY9 myxococcus	940	6	0.7	153	16	09PAN6	09PAN6 xyella fas
868	6	0.7	140	4	09P0F6	09P0F6 homo sapien	941	6	0.7	153	16	09JY03	09JY03 neisseria m
869	6	0.7	140	5	P91191	P91191 caenorhabdi	942	6	0.7	153	16	09JUP4	09JUP4 neisseria m
870	6	0.7	140	11	09D7H1	09D7H1 mus muscul	943	6	0.7	153	16	09J1I0	09J1I0 staphylococ
871	6	0.7	140	16	08VJEG	08VJEG pyrococcus	944	6	0.7	153	16	08RPM7	08RPM7 fusobacteri
872	6	0.7	140	17	059212	059212 pyrococcus	945	6	0.7	153	17	09HRP5	09HRP5 halobacteri
873	6	0.7	141	2	047941	047941 streptococ	946	6	0.7	153	17	09HKT6	09HKT6 thermoplas
874	6	0.7	141	9	09MBK7	09MBK7 bacterioph	947	6	0.7	153	17	08TUD5	08TUD5 methanosarc
875	6	0.7	141	12	072162	072162 tobacco rat	948	6	0.7	154	5	03VM41	03VM41 drosophila
876	6	0.7	141	16	09RVS8	09RVS8 delinococcus	949	6	0.7	154	8	032248	032248 gigartina a
877	6	0.7	141	16	08XLV9	08XLV9 clostridium	950	6	0.7	154	10	082011	082011 lycopersico
878	6	0.7	142	2	066363	066363 unidentified	951	6	0.7	154	10	082012	082012 lycopersico
879	6	0.7	142	2	070079	070079 unidentified	952	6	0.7	154	12	099A06	099A06 tt virus, a
880	6	0.7	142	10	09LV71	09LV71 arabisdopsis	953	6	0.7	154	16	091AT0	091AT0 pseudomonas
881	6	0.7	142	10	09FHC3	09FHC3 arabisdopsis	954	6	0.7	154	16	09HTM6	09HTM6 pseudomonas
882	6	0.7	142	12	09DKP6	09DKP6 hepatitis b	955	6	0.7	154	16	08ZKE6	08ZKE6 salmonella
883	6	0.7	142	16	08Z349	08Z349 salmonella	956	6	0.7	154	16	08X1P8	08X1P8 anabena sp
884	6	0.7	142	16	09K3K7	09K3K7 streptomyce	957	6	0.7	154	16	08YXYP	08YXYP salmonella
885	6	0.7	142	17	08U2E2	08U2E2 pyrococcus	958	6	0.7	154	16	09X034	09X034 streptomyce
886	6	0.7	143	4	08WVNS	08WVNS homo sapien	959	6	0.7	154	17	09HJK2	09HJK2 thermoplas
887	6	0.7	143	5	0931I1	0931I1 anophelies g	960	6	0.7	154	17	097W40	097W40 sulfolobus
888	6	0.7	143	5	020125	020125 caenorhabdi	961	6	0.7	155	5	09V262	09V262 aplysia cal
889	6	0.7	143	9	08SCP6	08SCP6 pseudomonas	962	6	0.7	155	5	09SVY9	09SVY9 drosophila
890	6	0.7	143	10	09SPH2	09SPH2 gossypium h	963	6	0.7	155	11	09CRN0	09CRN0 mus muscul
891	6	0.7	143	12	091N38	091N38 hepatitis b	964	6	0.7	155	16	09KXK7	09KXK7 streptomyce
892	6	0.7	143	12	091N37	091N37 hepatitis b	965	6	0.7	156	2	P96135	P96135 thermus the

966	6	0.7	156	2	0930W8	093qW8 pseudomonas
967	6	0.7	156	5	0227S7	0227S7 caenorhabdi
968	6	0.7	156	9	09VZNO	09vzn0 drosophila
969	6	0.7	156	9	08W6L2	08w6l2 sinorhizobi
970	6	0.7	156	10	08RMG7	08rmg7 arabidopsis
971	6	0.7	156	11	08VDG0	08vdg0 mus musculu
972	6	0.7	156	12	09WTR89	09wtr89 tt.virus. o
973	6	0.7	156	16	09JRS9	09jrs9 neisseria m
974	6	0.7	156	17	09YBZ9	09ybz9 aeropyrum p
975	6	0.7	157	2	045512	045512 bacillus su
976	6	0.7	157	2	09SL6	09sl6 streptococc
977	6	0.7	157	10	09LS13	09ls13 arabidopsis
978	6	0.7	157	16	08XRT5	08xrt5 ralsonia s
979	6	0.7	157	17	09H010	09h010 halobacteri
980	6	0.7	158	2	09F901	09f901 comanonas s
981	6	0.7	158	10	09M5F1	09m5f1 medicago sa
982	6	0.7	158	10	09C7P9	09c7p9 arabidopsis
983	6	0.7	158	16	09A0F1	09a0f1 streptococc
984	6	0.7	158	16	097PK5	097pk5 streptococc
985	6	0.7	159	2	049778	049778 mycobacteri
986	6	0.7	159	2	08VM06	08vm06 streptococc
987	6	0.7	159	10	09SX29	09sx29 arabidopsis
988	6	0.7	159	16	09RUC5	09ruc5 deinococcus
989	6	0.7	159	16	09KSH8	09ksh8 vibrio chol
990	6	0.7	159	16	09A1G2	09a1g2 streptococc
991	6	0.7	159	16	08YPC8	08ypc8 anabaena sp
992	6	0.7	160	2	030595	030595 escherichia
993	6	0.7	160	2	093M08	093m08 xanthomonas
994	6	0.7	160	3	099139	099139 ustilago sc
995	6	0.7	160	3	P78566	P78566 agaricus bl
996	6	0.7	160	16	031590	031590 bacillus su
997	6	0.7	161	2	030617	030617 myxococcus
998	6	0.7	161	2	093AN7	093an7 pseudomonas
999	6	0.7	161	3	014318	014318 schizosacch
1000	6	0.7	161	12	09E997	09e997 hepatitis b

ALIGNMENTS

RESULT 1
08ZB95 PRELIMINARY: PRT: 578 AA.

AC 08ZB95;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
GN Putative exported protein.

OS Versinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
NCBI_TaxID=632;
OX NCBI_TaxID=632;

RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RA MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Rarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;
RT "Genome sequence of *Yersinia pestis*, the causative agent of plague."
RL Nature 413:523-527(2001).

DR EMBL: AJ414157; CAC92753.1; -;
DR InterPro: IPR000566; Lipoclin_cytfabp.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.

SEQUENCE 578 AA: 64405 MW; 69AD3780BDC20F1 CRC64;

Query Match 1.4%: Score 13; DB 16; Length 578;

Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 880 GAGVGRWASPVG 892
Db 538 GAGVGRWASPVG 550

RESULT 2

08XEL5 PRELIMINARY: PRT: 577 AA.

AC 08XEL5;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE Putative outer membrane protein (Putative exported protein).
GN YTFM OR STM4409 OR STM4768.
OS Salmonella typhimurium, and

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
NCBI_TaxID=602, 601;
OX NCBI_TaxID=602, 601;

RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RA MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Portmullik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flora L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
LT2."

RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.

RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahina M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Fellwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsels K.,
RA Krogh A., Larsen T.S., Leather S., Hien T.T., Holroyd S., Jagsels K.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant *Salmonella*
enterica serovar Typhi CT18."

RL Nature 413:848-852(2001).
DR EMBL: AE008906; AL23229.1; -;
DR InterPro: IPR00184; Bac_surfA_D15.
DR InterPro: IPR000566; Lipoclin_cytfabp.
DR Pfam: PF01103; Bac_surface_Ag_1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.

SEQUENCE 577 AA: 64765 MW; FBA4190173BP18C CRC64;

Query Match 1.2%: Score 11; DB 16; Length 577;
Best Local Similarity 100.0%; Pred. No. 0.077;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 882 GGVGRWASPVG 892
Db 539 GGVGRWASPVG 549

RESULT 3

091001 PRELIMINARY: PRT: 579 AA.

AC 091001;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)

DT 01-OCT-2001 (TREMblrel. 18, last annotation update)
 DE Hypothetical protein PA2543.
 GN PA2543.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_Taxid=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stoyer C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 Ralzer J., Sailer M.H., Hancock R.E.W., Lory S., Olsen M.V.,
 RT *Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen.*;
 RL Nature 406:959-964(2000).
 DR EMBL; AE004682; AAG05931.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 579 AA; 63739 MW; 23467005C836FBF CRC64;

Query Match 1.18; Score 10; DB 16; Length 579;
 Best Local Similarity 100.0%; Pred. No. 0.81;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 809 LRFAGGDOS 818
 |||||
 DB 470 LRFAGGDOS 479

RESULT 4

ID 084951 PRELIMINARY; PRT; 260 AA.
 AC 084951;
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
 DE Ssef (secretion system effector).
 GN Ssef OR STM1404.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_Taxid=602;
 RN [1]
 RP SEQUENCE OF 253-260 FROM N.A.
 RC STRAIN-SL1344;
 RX MEDLINE=97451029; PubMed=9302299;
 RA Valdivia R.H., Falkow S.;
 RT *Fluorescence-based isolation of bacterial genes expressed within host
 cells.*;
 RL Science 277:2007-2011(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SL1344;
 RX MEDLINE=99000133; PubMed=9786194;
 RA Cifillio D.M., Valdivia R.H., Monack D.M., Falkow S.;
 RT *Macrophage-dependent induction of the Salmonella pathogenicity island
 2 type III secretion system and its role in intracellular survival.*;
 RL Mol. Microbiol. 30:175-188(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SL1344;
 RX Submitted (Jul-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA Mclelland M., Sanderson K.E., Spieth J., Clifton S.W., Lattelle P.,
 Courtney L., Porwollik S., All J., Dante M., Du F., Hou S., Layman D.,

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT *Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.*;
 RL Nature 413:852-856(2001).
 DR EMBL; AF020808; AAC28885.1; -;
 DR EMBL; AE008761; AAL20328.1; -;
 KW Complete proteome.
 SQ SEQUENCE 260 AA; 26725 MW; 0890B23F4FC0DE78 CRC64;

Query Match 1.08; Score 9; DB 16; Length 260;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 PVLTPEOI 66
 |||||
 DB 43 PVLTPEOI 51

RESULT 5

ID 09R802 PRELIMINARY; PRT; 262 AA.
 AC 09R802;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, last annotation update)
 DE Ssef.
 GN Ssef.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_Taxid=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2;
 RX MEDLINE=99000132; PubMed=9786193;
 RA Hensel M., Shea J.E., Waterman R., Mundy R., Nikolaus T., Banks G.,
 RA Vazquez-Torres A., Gleason C., Fang F.C., Holden D.W.;
 RT *Genes encoding putative effector proteins of the type III secretion
 system of Salmonella pathogenicity island 2 are required for bacterial
 virulence and proliferation in macrophages.*;
 RL Mol. Microbiol. 30:163-174(1998).
 DR EMBL; AJ224892; CAA12191.1; -;
 SQ SEQUENCE 262 AA; 26893 MW; 62A27ED053DDCEA CRC64;

Query Match 1.08; Score 9; DB 2; Length 262;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 PVLTPEOI 66
 |||||
 DB 43 PVLTPEOI 51

RESULT 6

ID 09CK26 PRELIMINARY; PRT; 586 AA.
 AC 09CK26;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
 DE Hypothetical protein PM1809.
 GN PM1809.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_Taxid=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;

```

RT      "Complete genomic sequence of Pasteurella multocida pm70."
RL      Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR      EMBL: AE006218; AKK03893.1; -.
DR      InterPro: IPR000184; Bac_SurfAg_D15.
DR      InterPro: IPR001092; HLH_Basic.
DR      Pfam: PF01103; Bac_surfAc_Ag; 1.
DR      ProSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 586 AA; 67106 MW; 00895174DED58283 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 16; Length 586;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      884 GVRMASPVC 892
        |||||||
Db      549 GVRMASPVC 557

RESULT 7
O9RPH7  PRELIMINARY; PRT; 1085 AA.
AC      O9RPH7;
DT      01-MAY-2000 (TREMBlrel. 13, Created)
DT      01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT      01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE      Rec.
DE      RECC.
GN      Mycobacterium smegmatis.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1772;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=99412429; PubMed=10481025;
RX      Griffin IV T.J., Parsons L., Leshchiner A.E., Devost J.,
RA      Derbyshire K.M., Grindley N.D.;
RT      "In vitro transposition of tm552: a tool for DNA sequencing and
RT      mutagenesis";
RL      Nucleic Acids Res. 27:3859-3865(1999).
DR      EMBL: AF157643; AAD46807.1; -.
SQ      SEQUENCE 1085 AA; 120352 MW; F4D9A81AA6A277C4 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 2; Length 1085;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      236 QTALVAAA 244
        |||||||
Db      753 QTALVAAA 761

RESULT 8
P95086  PRELIMINARY; PRT; 174 AA.
AC      P95086;
DT      01-MAY-1997 (TREMBlrel. 03, Created)
DT      01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT      01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE      Hypothetical protein RV3072C.
DE      RV3072C OR MT3157 OR MTCY22D7.09.
OS      Mycobacterium tuberculosis.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1773;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=H37RV;
RX      MEDLINE=98295987; PubMed=9634230;
RA      Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA      Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
RA      Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

```

```

RA      Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA      Hornsby T., Jagels K., Kiroh A., McLean J., Moule S., Murphy L.,
RA      Oliver S., Osborne J., Quail M.A., Rajandram M.A., Rogers J.,
RA      Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA      Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,
RT      "Deciphering the biology of Mycobacterium tuberculosis from the
RT      complete genome sequence.";
RL      Nature 393:537-544(1998).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CDC 1551 / OSHKOSH;
RA      Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA      Peterson J.F., Deboy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA      Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA      Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA      Bishai W.;
RT      "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT      laboratory strains.";
RL      Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL: Z83866; CAB06251.1; -.
DR      EMBL: AE007133; AAK47493.1; ALT_INT.
DR      TIGR: MT3157; -.
DR      Tuberculist; RV3072C; -.
DR      InterPro: IPR002103; Bac_Luciferase.
DR      Pfam: PF00296; bac_Luciferase; 1.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 174 AA; 17844 MW; 5073A9E87C0359D CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 174;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      895 RVDVATGV 902
        |||||||
Db      90 RVDVATGV 97

RESULT 9
O98JY3  PRELIMINARY; PRT; 235 AA.
AC      O98JY3;
DT      01-OCT-2001 (TREMBlrel. 18, Created)
DT      01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT      01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE      Hypothetical protein mlr1763.
DE      MLR1763.
OS      Rhizobium loti (Mesorhizobium loti).
OC      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC      Phyllobacteriaceae; Mesorhizobium.
OX      NCBI_TaxID=381;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=MAFF303099;
RX      MEDLINE=21082930; PubMed=11214968;
RA      Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA      Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA      Kishida Y., Kiyokawa C., Kohara M., Matsuno M., Matuno A.,
RA      Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA      Takeuchi C., Yamada M., Tabata S.;
RT      "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT      Mesorhizobium loti.";
RL      DNA Res. 7:331-338(2000).
DR      EMBL: AP002998; BAB49062.1; -.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 235 AA; 26023 MW; DB2C4A2DA6D19FC6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 235;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      238 ALVAAAV 245
        |||||||
Db      224 ALVAAAV 231

```



```

RESULT 10
098HK9
ID 098HK9 PRELIMINARY; PRT; 319 AA.
AC 098HK9;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Probable oxidoreductase.
GN ML2825.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shlimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003000; BAB49857.1; -
DR InterPro: IPR000683; GFO_IDH_MOCA.
DR InterPro: IPR004104; GFO_IDH_MOCA.C.
DR InterPro: IPR002016; Peroxidase.
DR Pfam; PF01408; GFO_IDH_MOCA.1.
DR Pfam; PF02894; GFO_IDH_MOCA.C; 1.
DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 319 AA; 35433 MW; A4F8C2E20977E352 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 319;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 VEVYGECA 287
Db 181 VEVYGECA 188

RESULT 11
0927W8
ID 0927W8 PRELIMINARY; PRT; 331 AA.
AC 0927W8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Mbl protein.
GN MBL OR LIN2669.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Chablit A., Chetouani F., Couve E., de Darvar A., Dehoux P.,
RA Domain E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkut G.,
RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,

```

```

RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596173; CAC97895.1; -
DR Listlist: LIN02669; -
DR InterPro: IPR001023; Hsp70.
DR InterPro: IPR004753; MreB_MrL.
DR PRINTS; PR01652; SHAPEPROTEIN.
DR Prodom; PD000089; Hsp70; 1.
DR TIGRFAMs; TIGR00904; mreB; 1.
KW Complete proteome.
SQ SEQUENCE 331 AA; 35581 MW; 1FB5D97D188CE7EA CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 331;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 839 GOVLAVGT 846
Db 38 GOVLAVGT 45

RESULT 12
08Y4C5
ID 08Y4C5 PRELIMINARY; PRT; 331 AA.
AC 08Y4C5;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Mbl protein.
GN MBL OR LMO2525.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Chablit A., Chetouani F., Couve E., de Darvar A., Dehoux P.,
RA Domain E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkut G.,
RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591983; CAD00603.1; -
DR Listlist: LMO02525; -
DR InterPro: IPR001023; Hsp70.
DR InterPro: IPR004753; MreB_MrL.
DR PRINTS; PR01652; SHAPEPROTEIN.
DR Prodom; PD000089; Hsp70; 1.
DR TIGRFAMs; TIGR00904; mreB; 1.
KW Complete proteome.
SQ SEQUENCE 331 AA; 35636 MW; 46A0FFC9CDE13825 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 331;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 839 GOVLAVGT 846
Db 38 GOVLAVGT 45

RESULT 13
092BL4

```

ID Q9ZBL4 PRELIMINARY; PRT; 340 AA.
 AC Q9ZBL4;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical 36.3 kDa protein (putative oxidoreductase).
 GN M0299 OR M0299.24
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Elgimeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies K.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus.";
 RL Nature 409:1007-1011(2001).
 DR EMBL: AL035159; CAA22708.1;
 DR EMBL: AL583918; CAC29807.1;
 DR Leproma: M0299;
 DR InterPro: IPR000927; D_aa-oxidase.
 DR Pfam: PF01266; DAO; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 340 AA; 36311 MW; E8D64FB7F817658B CRC64;

Query Match
 Best Local Similarity 0.9%; Score 8; DB 16; Length 340;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 884 GVRMASPV 891
 |||||
 DB 167 GVRMASPV 174

RESULT 14
 ID P72239 PRELIMINARY; PRT; 379 AA.
 AC P72239;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE CFA-beta-ketolactamase.
 GN CFA3.
 OS Pseudomonas syringae (pv. glyciniae).
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 NX NCBI_TaxID=318;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GLYCINA;
 RX MEDLINE=97149293; PubMed=8996103;
 RA Penfold C.N., Bender C.L., Turner J.G.;
 RT "Characterisation of genes involved in biosynthesis of coronafacic
 RT acid, the polyketide component of the phytotoxin coronatine.";
 RL Gene 183:167-173(1996).
 DR EMBL: U56980; AAB41300.1;
 DR HSSP: P73283; 1E5M.
 DR InterPro: IPR000794; ketoacyl-synt.
 DR Pfam: PF02801; ketoacyl-synt. 1.
 SQ SEQUENCE 379 AA; 39691 MW; 629745BCE98DA0B1 CRC64;

Query Match
 Best Local Similarity 0.9%; Score 8; DB 2; Length 379;
 Matches 8; Conservative 100.0%; Pred. No. 60;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 529 DAVSAVAR 536
 |||||
 DB 229 DAVSAVAR 236

RESULT 15
 ID Q9KYD1 PRELIMINARY; PRT; 383 AA.
 AC Q9KYD1;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative membrane protein.
 GN SC06899 OR SC182.05.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 NX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D.C., Harris D.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajadream M.A.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajadream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)."
 RL Nature 417:141-147(2002).
 DR EMBL: AL356812; CAB92561.1;
 DR InterPro: IPR000005; HTHARAC.
 DR PROSITE: PS00041; HTH ARAC FAMILY 1; UNKNOWN_1.
 SQ SEQUENCE 383 AA; 37020 MW; 9FFEP7E5827D4E CRC64;

Query Match
 Best Local Similarity 0.9%; Score 8; DB 16; Length 383;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 721 ALLGAVAV 728
 |||||
 DB 170 ALLGAVAV 177

RESULT 16
 ID Q8V5A8 PRELIMINARY; PRT; 452 AA.
 AC Q8V5A8;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TReMBLrel. 20, last sequence update)
 DE 01-MAR-2002 (TReMBLrel. 20, last annotation update)
 GN Nonstructural protein 2.
 OS NS2.
 OC Casphalia extranea densovirus.
 OC Viruses: ssDNA viruses; Parvoviridae; Densovirinae; Densovirus.
 NC NCBL_TaxID=180586;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tjissen P., Fediere G., Li Y., Zadori Z., Szelei J.;
 RT "Genome organization of Casphalia extranea densovirus (CedNV), a new
 RT iteravirus";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF375296; AAL56545.1; -
 SQ SEQUENCE 452 AA; 52387 MW; B0FBA9129C148C7 CRC64;
 QY 115 TRPISLEE 122
 DB 180 TRPISLEE 187

Query Match
 Best Local Similarity 100.0%; Score 8; DB 12; Length 452;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 17
 ID 086499 PRELIMINARY; PRT; 469 AA.
 AC 086499;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, last sequence update)
 DE 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE Hypothetical protein SC06331.
 GN SC06331 OR SC10H5.07.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 NC NCBL_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Krieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL; AL031232; CAA20279.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 469 AA; 52166 MW; A9D619820B16389A CRC64;

Query Match
 Best Local Similarity 100.0%; Score 8; DB 16; Length 469;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 PVALAAYL 21
 DB 420 PVALAAYL 427

RESULT 18
 ID 076799 PRELIMINARY; PRT; 506 AA.
 AC 076799;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)

DE Microsomal cytochrome P450.
 GN CYP9B3.
 OS Drosophila mettleri (Fruit fly).
 OS OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBL_TaxID=7228;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Danielson P.B., Reeves S.A., Fogleman J.C.;
 RT "Isolation of a novel cytochrome P450 (CYP9B3) from Drosophila
 RT mettleri";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AF083945; AUC33297.1; -
 DR FlyBase; FBgn0025418; Dmel:CYP9B3.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450; 1.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 506 AA; 5855 MW; 246BC0510F6901D0 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 8; DB 5; Length 506;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 LSNDLIAT 417
 DB 185 LSNDLIAT 192

RESULT 19
 ID 087139 PRELIMINARY; PRT; 508 AA.
 AC 087139;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE Phycocyanin alpha phycocyanobilin lyase related protein.
 GN MA4317.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 NC NCBL_TaxID=2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE-21929760; PubMed-11932238;
 RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Galvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umeyan L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity";
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AE011149; AAM07661.1; -
 KW Lyase; Complete proteome.
 SQ SEQUENCE 508 AA; 55620 MW; A07C705CBCA80278 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 8; DB 17; Length 508;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL5 411
 DB 264 LQAVRAL5 271

```

RESULT 20
O9KRP31
ID O9KRP31 PRELIMINARY: PRT: 582 AA.
AC O9KRP31:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein VC2548.
GN VC2548.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Elsen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nieman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RT Nature 406:477-483(2000).
RL EMBL: AE004323; AAF95689.1; -.
DR TIGR: VC2548; -.
DR InterPro: IPR000184; Bac_surfAg_D15.
DR InterPro: IPR000566; Lipocin_cytfABP.
DR Pfam: PF01103; Bac_surfAg; 1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 582 AA; 65644 MW; E96946B85F1C426 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 582;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 809 LRFPGAGD 816
DB 473 LRFPGAGD 480

RESULT 21
O8ZSS3
ID O8ZSS3 PRELIMINARY: PRT: 731 AA.
AC O8ZSS3:
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein A117659.
GN A117659.
OS Anabaena sp. (strain PCC 7120).
OC Plasmid PCC7120beta.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kurlitz T., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL: AF003602; BAB7302.1; -.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 731 AA; 83004 MW; DEAA8AA4F02DE54D5 CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 731;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 212 IKALEDI 219
DB 116 IKALEDI 123

RESULT 22
O9HK17
ID O9HK17 PRELIMINARY: PRT: 1047 AA.
AC O9HK17:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Probable cardamoyl-phosphate synthase, large subunit.
GN TA0791.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumelster W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum.";
RT Nature 407:508-513(2000).
RL EMBL: AL445065; CAC11922.1; -.
DR HSSP: P00968; LJDB.
DR InterPro: IPR000901; CPsase.
DR InterPro: IPR004362; MGS_like.
DR Pfam: PF00289; CPsase_L-chain; 2.
DR Pfam: PF02786; CPsase_L-D2; 1.
DR Pfam: PF02787; CPsase_L-D3; 1.
DR Pfam: PF02142; MGS; 1.
DR PRINTS: PR00098; CPsase.
DR PROSITE: PS00867; CPsase_2; UNKNOWN_2.
DR Complete proteome.
SQ SEQUENCE 1047 AA; 116851 MW; E062CAD131746864 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 17; Length 1047;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 626 KGKLYAT 633
DB 941 KGKLYAT 948

RESULT 23
O9VKH2
ID O9VKH2 PRELIMINARY: PRT: 1373 AA.
AC O9VKH2:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE SALM protein.
GN SALM OR CG6464.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

```

RA Amaratides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,
RA Abill J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikva D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dopp L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Sprelling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003632; AAF53097.1; -.
DR HSSP: P15822; 1BBO.
DR Flybase: FBgn0004579; salm.
DR InterPro: IPR000822; znf.C2H2.
DR Pfam: PF00096; zf-C2H2; 7.
DR SMART: SM00355; znf.C2H2; 7.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.
KM DNA-binding: Metal-binding; Zinc-finger.
SQ SEQUENCE 1373 AA; 151014 MW; A6D857870F645ACC CRC64;

Query Match 0.9%; Score 8; DB 5; Length 1373;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 61 LTRPQIOA 68
DB 913 LTRPQIOA 920
RESULT 24
Q48697 PRELIMINARY; PRT; 1417 AA.
AC Q48697;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE F316.24 protein.
GN F316.24.
GN Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-CV, COLUMBIA;
RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,
RA Au M., Ataulo R., Buehler E., Dewar K., Feng J., Kim C., Li Y.,
RA Oji O., Osborne B.L., Shinn P., Sun H., Toriumi M., Vysotskaya V.S.,
RA Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC002396; AAC00591.1; -.
DR InterPro: IPR003169; GYF.
DR Pfam: PF02213; GYF; 1.
DR SMART: SM00444; GYF; 1.
SQ SEQUENCE 1417 AA; 157750 MW; 5B08FBAC1F48334 CRC64;

Query Match 0.9%; Score 8; DB 10; Length 1417;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 387 RELLEOLL 394
DB 843 RELLEOLL 850

RESULT 25
Q45283 PRELIMINARY; PRT; 33 AA.
AC Q45283;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE Alpha-amylase precursor (Fragment).
GN AMYL.
OC Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91092499; PubMed=2265757;
RA Jorgensen P.L., Hansen C.K., Poulsen G.B., Diderichsen B.;
RT "In vivo genetic engineering: homologous recombination as a tool for
RT plasmid construction.";
RL Gene 96:37-41(1990).
DR EMBL: M62637; AAA2232.1; -.
DR SIGNAL.
KT SIGNAL. 1 29
FT CHAIN 30 >33
FT NON_TER 33
FT POTENTIAL.
FT POTENTIAL.
SQ SEQUENCE 33 AA; 3711 MW; 9CA3D49C5F5C36F7 CRC64;

Query Match 0.8%; Score 7; DB 2; Length 33;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 166 KRLYARL 172
DB 5 KRLYARL 11
RESULT 26
Q8W6C9 PRELIMINARY; PRT; 62 AA.
AC Q8W6C9;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Menda.
GN ORF62.
GN Viruslike phage VSK.
OC Viruses; ssDNA viruses; Inoviridae; Inovirus.
OX NCBI_TaxID=181604;
RN [1]
RP SEQUENCE FROM N.A.
RA Basu N., Kar S., Ghosh R.K.;
RT "Molecular analysis of filamentous phage VSK of Vibrio cholerae 0139;

```

RT A possible clue to genetic transmission."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF453500; AAL49745.1;
SQ SEQUENCE 62 AA; 7221 MW; 56813397E7BCBDF CRC64;

Query Match
Best Local Similarity 0.8%; Score 7; DB 9; Length 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 888 ASPVGOV 894
    |||||
Db 37 ASPVGOV 43

RESULT 27
O8M542 PRELIMINARY; PRT; 66 AA.
AC O8M542;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-JUN-2002 (TREMBlrel. 20, Last sequence update)
DE Dehydrin (Fragment).
GN DHN2.
OS Retama raetam.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Retama.
OX NCBI_TaxID=49837;
RN [1]
RP SEQUENCE FROM N.A.
RA Pnuel L., Mitter R.;
RT "Dehydrin in retama raetam.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF439276; AAL32035.1;
DR InterPro: IPR000167; Dehydrin.
DR Pfam: PF00257; dehydrin; 1.
FT NON_TER
FT NON_TER
SQ SEQUENCE 66 AA; 7354 MW; 857C8C3DE22912P6 CRC64;

Query Match
Best Local Similarity 0.8%; Score 7; DB 10; Length 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 150 EVVVPPT 156
    |||||
Db 43 EVVVPPT 49

RESULT 28
O9PD36 PRELIMINARY; PRT; 66 AA.
AC O9PD36;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2002 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein Xf1543.
GN Xf1543.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9ASC.
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Agencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.B.A., Carraro D.M., Carter H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,

```

```

RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsumura A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Pelxoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL: AE003984; AAP84352.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 66 AA; 6871 MW; 8DA1217E63550F41 CRC64;

Query Match
Best Local Similarity 0.8%; Score 7; DB 16; Length 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 809 LRFPAAG 815
    |||||
Db 11 LRFPAAG 17

RESULT 29
P74009 PRELIMINARY; PRT; 72 AA.
AC P74009;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein ssr2201.
GN SSR2201.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasaoka S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nairuo K., Okumura S.,
RA Shlomo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL: D90911; BAA18080.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 72 AA; 8400 MW; C66AE11FE431E0EF CRC64;

Query Match
Best Local Similarity 0.8%; Score 7; DB 16; Length 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 216 LEDITQE 222
    |||||
Db 54 LEDITQE 60

RESULT 30
O8XG35

```

```

ID 08XGJ5 PRELIMINARY; PRT: 82 AA.
AC 08XGJ5;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hydrogenase-2 operon protein (Hydrogenase-2 component protein).
GN HtgB OR STM3143 OR STY3314.
OS Salmonella typhimurium, and
OC Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA Mclelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Portnoylik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2."
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Kirogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR EMBL: AE008844; AAL22017.1; -
DR EMBL: AL672727; CAD02975.1; -
DR InterPro: IPR001109; HUPF_HyPC.
DR Pfam: PF01455; HUPF_HyPC.
DR PRINTS: PR00445; HUPFHYP.
DR PRODOM: PD003112; HUPF_HyPC.1.
DR TIGRFAMs: TIGR00074; hupc_hupc.1.
DR PROSITE: PS01097; HUPF_HyPC.1.
KW Complete proteome.
SQ SEQUENCE 82 AA; 8850 MW; 381D32AB941860E7 CRC64;

Query Match 0.8%; Score 7; DB 16; Length 82;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 839 GOVLAVG 845
Db 7 GOVLAVG 13

RESULT 31
OBTREI PRELIMINARY; PRT: 84 AA.
AC OBTREI;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Molybdenum-pterin binding protein.
GN MA139.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932228;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArelano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirelli A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL: AE010791; AAM04658.1; -
KW Complete proteome.
SQ SEQUENCE 84 AA; 8963 MW; 55C35F09C61143C3 CRC64;

```

```

Query Match 0.8%; Score 7; DB 17; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 MNVTEIV 428
Db 35 MNVTEIV 41

```

```

RESULT 32
O9PVI8 PRELIMINARY; PRT: 85 AA.
AC O9PVI8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Regulator of G-protein signalling 12 (Frigment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Schiffl M., Jordan J.D., Diverse-Pierluisi M.;
RT "Mechanisms of desensitization of G-activated tyrosine kinase
pathways."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC 1- SIMILARITY: CONTAINS 1 RGS DOMAIN.
DR EMBL: AF090086; AAF00029.1; -
DR HSSP: P49799; IAGR.
DR InterPro: IPR000342; Regl_Gprotein.
DR Pfam: PF00615; RGS.1.
DR PRINTS: PR01301; RGSPTROTEIN.
DR PRODOM: PD001580; Reg_of_prg.1.
DR SMART: SM00315; RGS.1.
DR PROSITE: PS50132; RGS.1.
FT NON_TER 1
FT TER 85
SQ SEQUENCE 85 AA; 10108 MW; A96F32B2BBA97D59 CRC64;

```

```

Query Match 0.8%; Score 7; DB 13; Length 85;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 NHVPAHD 46
Db 15 NHVPAHD 21

```

RESULT 33

09CLV3
ID 09CLV3 PRELIMINARY: PRT: 87 AA.
AC 09CLV3
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical protein PM1099.
GN PM1099.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70:
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L., Paustian M.L., Whitlam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida PM70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL: AE006151; AAK03183.1;-
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 87 AA; 9829 MW; 5960BFA5BAAC474 CRC64;

Query Match 0.8%; Score 7; DB 16; Length 87;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSKRVLF 7
DB 1 MSKRVLF 7

RESULT 34
082102
ID 082102 PRELIMINARY: PRT: 88 AA.
AC 082102;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Putative flagellar regulatory protein.
GN YPO0720.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Tildall R.W., Holden M.T.G.,
RA Prentice M.B., Sebalha M., James K.D., Churcher C., Mungall K.L.,
RA Chillingworth T., Cronin N., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL: AA414144; CAC89571.1;-
KW Flagella; Hypothetical protein; Complete proteome.
SQ SEQUENCE 88 AA; 9503 MW; DCCB35E838168073 CRC64;

Query Match 0.8%; Score 7; DB 16; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 533 AVARAIL 539
DB 77 AVARAIL 83

RESULT 35
09JSVO

ID 09JSVO PRELIMINARY: PRT: 91 AA.
AC 09JSVO;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical protein NMA2123.
GN NMA2123.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=656599;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Fellwell T., Hamlin N., Holtroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491.";
RL Nature 404:502-506(2000).
DR EMBL: AL162758; CAB85336.1;-
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 91 AA; 10240 MW; 6327497E655F2DC0 CRC64;

Query Match 0.8%; Score 7; DB 16; Length 91;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 LTPEDIQ 67
DB 8 LTPEDIQ 14

RESULT 36
098552
ID 098552 PRELIMINARY: PRT: 95 AA.
AC 098552;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE A502L protein.
GN A502L.
OS Paramyxium bursaria chlorella virus 1 (PBCV-1).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.
OX NCBI_TaxID=10506;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96400190; PubMed=8806566;
RA Kutish G.F., Li Y., Lu Z., Furuta M., Rock D.L., Van Etten J.L.;
RT "Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: map
RT positions 182 to 258.";
RL Virology 223:303-317(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20013326; PubMed=10544099;
RA Kaiser A., Vollmert M., Tholl D., Graves M.V., Gurnon J.R., Xing W.,
RA Liseac A.D., Nickerson K.W., Van Etten J.L.;
RT "Chlorella virus PBCV-1 encodes a functional homospemidine
RT synthase.";
RL Virology 263:254-262(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20478054; PubMed=11021991;
RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;
RT "Characterization of a beta-1,3-glucanase encoded by chlorella virus
RT PBCV-1.";
RL Virology 276:27-36(2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Van Etten J.L.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Van Etten J.L.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RA Graves M.V., Van Etten J.L.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE FROM N.A.
 RA Graves M.V., Van Etten J.L.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [9]
 RP SEQUENCE FROM N.A.
 RA Curron J.R., Graves M.V., Van Etten J.L.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U42580; AAC96869.1; -
 SQ SEQUENCE 95 AA; 11069 MW; 0811C752B1618A80 CRC64;

Query Match 0.88; Score 7; DB 12; Length 95;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 817 OSTRGYA 823
 DB 47 OSTRGYA 53

RESULT 37
 ID 09UP45 PRELIMINARY; PRT; 98 AA.
 AC 09UP45;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE Type XIII collagen (Fragment).
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99357014; PubMed=10429945;
 RA Kvist A.P., Latvanlehto A., Sund M., Horelli-Kuitunen N., Rehn M.,
 RA Palotie A., Belter D., Pihlajaniemi T.;
 RT "Complete exon-intron organization and chromosomal location of the
 RT gene for mouse type XIII collagen (coll13a1) and comparison with its
 RT human homologue."
 RL Matrix Biol. 18:261-274(1999).
 DR EMBL: AF071009; AAD48082.1; -
 KW Collagen.
 FT NON_TER
 SQ SEQUENCE 98 AA; 10228 MW; 93020572713BEE29 CRC64;

Query Match 0.88; Score 7; DB 4; Length 98;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 238 ALVAAA 244
 DB 27 ALVAAA 33

RESULT 38
 ID 096YV4 PRELIMINARY; PRT; 105 AA.
 AC 096YV4;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical protein str2073.
 GN str2073.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=111955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JCM 10545 / 7;
 RX PubMed=11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 RA Yoshizawa T., Tanaka T., Kudo Y., Yamazaki J., Kushiya N., Oguchi A.,
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
 RA Oshima T., Kikuchi H.;
 RT "Complete genome sequence of an aerobic thermophilic
 RT Crenarchaeon, Sulfolobus tokodaii strain 7";
 RL DNA Res. 8:123-140(2001).
 DR EMBL: AP000988; BAB67172.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 105 AA; 12168 MW; C2FC9E804E1AC83C CRC64;

Query Match 0.88; Score 7; DB 17; Length 105;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 249 DIDSLII 255
 DB 74 DIDSLII 80

RESULT 39
 ID 092GX1 PRELIMINARY; PRT; 119 AA.
 AC 092GX1;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 50S ribosomal protein L22.
 GN RPLV OR RCL001.
 OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MALISH 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii".
 RL Science 293:2093-2098(2001).
 DR EMBL: AEO08652; AAL03539.1; -
 DR InterPro: IPR001063; Ribosomal_L22.
 DR Pfam: PF00237; Ribosomal_L22; 1.
 DR ProDom: PD001032; Ribosomal_L22; 1.
 DR TIGRFAMs: TIGR01044; rplV_bact; 1.
 DR PROSITE: PS00464; RIBOSOMAL_L22; UNKNOWN_1.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 119 AA; 13253 MW; 5BE636CA907C7DA8 CRC64;

Query Match 0.88; Score 7; DB 16; Length 119;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 491 KLNLYAA 497
 DB 22 KLNLYAA 28

RESULT 40

```

ID 028742 PRELIMINARY; PRT; 119 AA.
AC 028742;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein AF1530.
GN AF1530.
OS Archaeoglobus fulgidus.
OC Archaea: Euryarchaeota: Archaeoglobi: Archaeoglobales:
OC Archaeoglobaceae: Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kinkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Relch C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL: AE000997; AAB89718.1; -.
DR TIGR: AF1530; -.
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 119 AA; 14025 MW; 7F75DA933E80C15C CRC64;

```

Query March 0.8%; Score 7; DB 17; Length 119;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LTPEDIQ 67

Db 102 LTPEDIQ 108

Search completed: April 28, 2003, 16:35:35
Job time : 159 secs

CC MC391 (ATCC 43617) BASB081 mature protein. BASB081 is related by
 CC amino acid sequence homology to *Neisseria meningitidis* omp85 outer
 CC membrane protein. The sequence was deduced from PCR-amplified DNA
 CC (see AAA0537), and shows 99.9% homology to the mature region of a
 CC BASB081 gene translation product (see AAY95987). The invention
 CC provides BASB081 polypeptides, polynucleotides, expression vectors,
 CC host cells, and a process for producing a BASB081 polypeptide. Also
 CC provided are vaccine compositions comprising a BASB081 polypeptide
 CC or polynucleotide, and optionally at least 1 other *M. catarrhalis*
 CC antigen. A method for diagnosing a *M. catarrhalis* infection
 CC involves identifying a BASB081 polypeptide, or an antibody that is
 CC immunospecific for it, in a sample. A therapeutic composition
 CC useful in treating *M. catarrhalis* diseases in humans comprises an
 CC antibody directed against a BASB081 polypeptide. The disease can
 CC be a bacterial infection, e.g. otitis media in infants and children,
 CC pneumonia in elderlies, sinusitis, nosocomial infections and
 CC invasive diseases, chronic otitis media with hearing loss, fluid
 CC accumulation in the middle ear, auditive nerve damage, delayed
 CC speech learning, upper respiratory tract infection, and inflammation
 CC of the middle ear.

CC Sequence 889 AA:

Query Match. 96.9%: Score 4581: DB 21: Length 889:

Best Local Similarity 99.9%: Pred. NO. 7e-307: Indels 0: Gaps 0:

Matches 888: Conservative 1: Mismatches 0: Indels 0: Gaps 0:

QY 31 QONNPANIINHVPADPAALNOAKAGNPVLLTPQIOARLNAAGLNAPQSQALDVAFD 90
 DB 1 QONNPANIINHVPADPAALNOAKAGNPVLLTPQIOARLNAAGLNAPQSQALDVAFD 60
 QY 91 DQSFISIGESPPGLGDMYSIETTPLSLEELFAQESTENGINDPIPEYQGEOPNSE 150
 DB 61 DQSFISIGESPPGLGDMYSIETTPLSLEELFAQESTENGINDPIPEYQGEOPNSE 120
 QY 151 VWPPTLEPEKPGILIKRLYARLFENDGVNKPRLAKFPQSSQSETSAIGSSHQKTEPYA 210
 DB 121 VWPPTLEPEKPGILIKRLYARLFENDGVNKPRLAKFPQSSQSETSAIGSSHQKTEPYA 180
 QY 211 NIKAALEDITQESAMDLNGSTIPRLQOTALVARAVGYDIDLSIIRNSIGEVYIIHDLG 270
 DB 181 NIKAALEDITQESAMDLNGSTIPRLQOTALVARAVGYDIDLSIIRNSIGEVYIIHDLG 240
 QY 271 EPVYIDYRAVEVRGEGADKAFVTVADEVPLLIGDVFHKGKYEKKNLINENASAEHGYD 330
 DB 241 EPVYIDYRAVEVRGEGADKAFVTVADEVPLLIGDVFHKGKYEKKNLINENASAEHGYD 300
 QY 331 GRMLDRSVDVILPONTADVSLIYDTGTYRFDVYVFFITDPKTNQLTDPDKLPVKREL 390
 DB 301 GRMLDRSVDVILPONTADVSLIYDTGTYRFDVYVFFITDPKTNQLTDPDKLPVKREL 360
 QY 391 EQLLTVMNGEAYNLQAVRALSNDLIATRYFMVNTETVFFPREOIONDOVSFEQSSSRT 450
 DB 361 EQLLTVMNGEAYNLQAVRALSNDLIATRYFMVNTETVFFPREOIONDOVSFEQSSSRT 420
 QY 451 EPAQVDESTLEPVYETVELFDGILMDISPIEFSASNLIDQKLNLAAKARHLVMPDDRV 510
 DB 421 EPAQVDESTLEPVYETVELFDGILMDISPIEFSASNLIDQKLNLAAKARHLVMPDDRV 480
 QY 511 LAINHDDGVNRSLIGRTSDAVSAAARAILPDESENEVIDLPERPALNKRKPAAVYQSKK 570
 DB 481 LAINHDDGVNRSLIGRTSDAVSAAARAILPDESENEVIDLPERPALNKRKPAAVYQSKK 540
 QY 571 VPLVYFVASDKPRDQGIIGLGMGSDPTGRLVTKFEHNLIINRDGYOAGAEELRLSEKKKGVL 630
 DB 541 VPLVYFVASDKPRDQGIIGLGMGSDPTGRLVTKFEHNLIINRDGYOAGAEELRLSEKKKGVL 600
 QY 631 YATKPLSHPLNDOLRATLGYQVEVFGHSTNGFDLSTRTLEHISRSIIIONGMWRTYSLR 690
 DB 601 YATKPLSHPLNDOLRATLGYQVEVFGHSTNGFDLSTRTLEHISRSIIIONGMWRTYSLR 660
 QY 691 YRLDKLKTQAPPEFTMODLPVDFVNGKPSQFALLAGVAVHKTVDNLVPMRGYQQRISLE 750

DB 661 YRLDKLKTQAPPEFTMODLPVDFVNGKPSQFALLAGVAVHKTVDNLVPMRGYQQRISLE 720
 QY 751 VGSSGLVSDAMMAIARAGISGYISFGDNAGSNRAHQMTGGIOAGYIMSDNFNVHYRLR 810
 DB 721 VGSSGLVSDAMMAIARAGISGYISFGDNAGSNRAHQMTGGIOAGYIMSDNFNVHYRLR 780
 QY 811 FFAGGDSIRGYAHDSLSPISDKGLTGGQYLAVTAETYNFPMKDLRLAVFGDIGNAV 870
 DB 781 FFAGGDSIRGYAHDSLSPISDKGLTGGQYLAVTAETYNFPMKDLRLAVFGDIGNAV 840
 QY 871 KGFMTDTKIGAGVGRMASPVGOVVDVATGVKKEGNPKLHFFIGTPE 919
 DB 841 KGFMTDTKIGAGVGRMASPVGOVVDVATGVKKEGNPKLHFFIGTPE 889

RESULT 3
 ID AAY95820 standard; Protein: 578 AA.
 AC AAY95820;
 DT 07-NOV-2000 (first entry)
 DE Haemophilus influenza strain Rd KW20 BASB067 protein.
 KW BASB067; outer membrane protein; antigen; vaccine; antibiotic;
 KW antibacterial; screening; infection; diagnosis; therapy.
 OS Haemophilus influenzae.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label_ Signal_peptide
 FT /label_ 23..578
 FT /label_ Mature_protein
 FT Domain 23..236
 FT /note= "N-terminal domain"
 FT Domain 237..578
 FT /note= "C-terminal domain"
 PN WO200047737-A1.
 XX
 PD 17-AUG-2000.
 PF 04-FEB-2000; 2000WO-EP00887.
 XX
 PR 09-FEB-1999; 99GB-0002880.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PI Ruelle J, Thonnard J;
 DR WPI: 2000-515059/46.
 DR N-PSDB; AAA50269.
 XX
 PS Claim 1; Page 78-80; 87pp; English.

The present sequence is that of outer membrane protein BASB067 of *Haemophilus influenzae* strain Rd KW20. BASB067 is a surface expressed protein that is recognised by the immune system. It shows 23% identity to the protective surface antigen D15 of *H. influenzae* and has a similar secondary structure. Its N-terminal domain is predicted to contain a mixture of alpha-helix and beta-strands, and could be used as a vaccine antigen. The C-terminal domain is predicted to form a beta-barrel composed of anti-parallel, amphipathic beta-strands. The external loops of the beta-barrels of integral outer membrane proteins frequently contain immunodominant B-cell epitopes, making the C-terminal domain of BASB067 a strong candidate vaccine antigen. The invention relates to recombinant materials and methods for the

CC production of BAsB067 polypeptides and polynucleotides, for use
CC especially in therapeutic and prophylactic vaccines. It also
CC relates to methods for using such polypeptides and polynucleotides
CC in the prevention and treatment of microbial diseases, in diagnostic
CC assays for detecting diseases associated with microbial infections,
CC and assays for detecting expression or activity of BAsB067
CC polypeptides or polynucleotides. Antibodies raised against
CC BAsB067 can be used to treat humans with H. influenzae disease.

SQ Sequence 578 AA;

Query Match	9.28;	Score 436.5;	DB 21;	Length 578;
-------------	-------	--------------	--------	-------------

Matches 142; Conservative 105; Mismatches 235; Indels 211; Gaps 14;

[illegible]

XX BAsR067: outer membrane protein; antigen; vaccine; antibiotic;
 KM antibacterial; screening; infection; diagnosis; therapy.
 KM
 XX
 OS Haemophilus influenzae.

Key	Location/Qualifiers
Peptide	1..20
Protein	/label= Signal-peptide 21..576
	/label= Mature-protein

PN WO200047737-A1

PD 17-AUG-2000.

PF 04-FEB-2000; 2000WO-EP00887.

PR 09-FEB-1999; 99GB-0002880.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Ruelle J, Thonnard J;

DR WPI; 2000-515059/46.

XX
DE

PT used for diagnosing and treating H. influenzae infections -

PS Claim 1; Page 81-82; 87pp; English

The present sequence is that of outer membrane protein BASH067 of Hemophilus influenzae non-typable (Nthi) strain 289. BASH067 is a surface expressed protein that is recognised by the immune system. It shows homology to the protective surface antigen D15 of H. influenzae and has a similar secondary structure. Its N-terminal domain is predicted to contain a mixture of alpha-helix and beta-strands, and could be used as a vaccine antigen. The C-terminal domain is predicted to form a beta-barrel composed of anti-parallel, amphipathic beta-strands. The external loops of the beta-barrels of integral outer membrane proteins frequently contain immunodominant B-cell epitopes, making the C-terminal domain of BASH067 a strong candidate vaccine antigen. The invention relates to recombinant materials and methods for the production of BASH067 polypeptides and polynucleotides, for use especially in therapeutic and prophylactic vaccines. It also relates to methods for using such polypeptides and polynucleotides in the prevention and treatment of microbial diseases, in diagnostic assays for detecting diseases associated with microbial infections, and assays for detecting expression or activity of BASH067 polypeptides or polynucleotides. Antibodies raised against the mature portion of this BASH067 polypeptide can be used to treat humans with H. influenzae disease.

50 Sequence 576 AA.

Query Match	9.08;	Score 427.5;	DB 21;	Length 576;
-------------	-------	--------------	--------	-------------

Matches 141; Conservative 106; Mismatches 235; Indels 211; Gaps 14;

DQ 243 RANVYDIDSLIRNS-IGEVNDVIHDL--GEFVYIDYRAVEKRGEGADKAFTVADEV 299

Db 71 RVGEVYESSVRFEKROGROGRDLLIAHTVPGEPTKLINGTGVQIEGGAQOENFDALRKNL 130

DQ 300 PLLIGDVHHGKAEETKKNLLENMSAHEGYDGDKMLDRSVUYIIPDNTADVSLIYDTGTQY 359

Dd 131 P-KEGLAVEHQTDYDYKTALSRILALNKGVEDGNEFKISRIELISPETHQAWRMFLFGSGRY 189

DQ 360 REPEVVEFTIDPRTNQTLTPDKLPVKRELLBOLLVYNMGEALINCAVARLSMDLIATRY 419

Dd 190 HYNATTF-----SHSQTRDVLNNILINKSDSPYLIMNLSLDTSDFSSSNW 235

```
QY 420 FNNVNTLIVPEREIOINDQVSEFOSSSSKTEPAQVDESTLEPIVETVELTDGILMDISP 479
|:|
Db 236 FSSV----- 239
QY 480 IEFASNLIDOKLNLVAAKARHLVMPDDRYLALNHDDGVNRSILGRISDAVSAVARAIL 539
|:|
Db 240 -----LVQPNV----- 246
QY 540 PDESENEVIDLPERTALANRKPADVYQSKKVPYVAVASDKPRDGOIGLGSGSDTGTRL 599
|:|
Db 247 -----HKSTVDVEITLLYPRKKNAMELGCAFADGCVHG 280
QY 600 VTKEFENLNRDQYQAQAEIRLSDKKGVK-LYATKPLSHPLNDQLRATLGYQGEVFGHS 658
|:|
Db 281 QIGETKPMINSRGLSRSLNLYLAPKQTLKATYMPPLKPLNLYYDFAVGWE-----GEK 336
QY 659 TNGFDLSTRTLEHEISRSIIONGGMNRTYSLRYRLDKLTKQAPPETQODLPVDVNGKPS 718
|:|
Db 337 EN--DTMTRVLTLSALRYNNNAHGMQYFGGLRTYDSF-TQAD----- 376
QY 719 OEALLAGVAHVKTVADNLVNPMPRGYR-----QRYSLSEVSSGLVSDANMA 763
|:|
Db 377 -----ITDKTL--LXPTVGFTRTLKSGSFATMGDQKITPDLKSRIMLSESSFI 425
QY 764 IARAGISGVYSEFGDNVAGSNRAHOMTGIQAGIYMSDNFNHVPYRLRPFAGDOSIRGYA 823
|:|
Db 426 KYQASSAMVRYAEN-----HRIVARAEIGYLHTKDIKPIPTLRFPAAGDSRVGCVG 478
QY 824 HDLSLPSIDSGYLTGGVLAIVGTAEVYEFMKDLRLAVFGDIGAVYKGFNDOKIGAGV 883
|:|
Db 479 YKKIAPKRNKGLKVGSRLLTSLENOYQYPPMMKAATFADSGLAUDNYTAKELRYGTGV 538
QY 884 GVRMASPVGVRVDVATVKEEGNPILKLFEEIG 916
|:|
Db 539 GVRMASPVGAIKFDIATPIRDKMSKNIQFYIG 571

RESULT 5
AAU03958
ID AAU03958 standard; Protein; 792 AA.
XX
AC AAU03958;
XX
DT 23-OCT-2001 (first entry)
XX
DE Neisseria meningitidis serogroup A antigenic protein #1.
XX
KM Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;
XX
KM bacterial infection; baculovirus; yeast.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT Protein /note= "Signal peptide"
FT Protein 22..792
FT Protein /note= "Mature N. meningitidis serogroup A antigen"

WO200138350-A2.
PD 31-MAY-2001.
PF 28-NOV-2000; 2000MO-IB01851.
XX
PR 29-NOV-1999; 99GB-0028197.
PR 09-MAR-2000; 2000GB-0005698.
XX
PA (CHIR-) CHIRON SPA.
PA (STAT-) STATENS INST FOLKEHELS.
PI Giuliani MM, Pizze M, Rappuoli R, Holst J.
DR WPI: 2001-381289/40.
```

```
DR N-PSDB: AAS07278.
XX
PT Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
PT gonorrhoeae, useful in the manufacture of a medicament for treating and
PT preventing Neisseria bacteria infection.
XX
PS Claim 1; Page 66-68; 92pp; English.
XX
CC The sequence represents a Neisseria meningitidis serogroup A 85 kDa
CC antigenic protein. Neisseria meningitidis colonises the pharynx, causing
CC meningitis and, occasionally, septicaemia in the absence of meningitis.
CC This antigenic protein is useful in the manufacture of a medicament for
CC treating or preventing infection due to Neisseria bacteria, such as
CC meningitis and septicaemia. It is also useful as a diagnostic reagent for
CC detecting the presence of Neisseria bacteria or antibodies raised against
CC Neisseria, and as a reagent for raising the antibodies. The Neisseria
CC nucleotide sequences can be expressed in a variety of different
CC expression systems, for example, mammalian cells, baculoviruses, plants,
CC bacteria and yeast.
CC Note: There are two versions of this sequence displayed in the
CC specification (see AAU04451).
XX
SQ Sequence 792 AA;
XX
Query Match 5.6%; Score 265.5; DB 22; Length 792;
Best Local Similarity 20.9%; Pred. No. 1.1e-09;
Matches 192; Conservative 119; Mismatches 352; Indels 255; Gaps 39;
QY 116 TPLSLEELFAOESTREMGINPND-----YIPYQGEQPNSEFVVPPLPEPKLIRLY 169
|:|
Db 16 SPLAFADFTIODIRVEGLQRTPESTVFNYLPVKGDYTN-----DTGSAIKSLY 66
QY 170 ARLENDGVNKP---RLKAKFYQSSQSGENTSATISSHOKTEPYANIKALED--ITQESA 224
|:|
Db 67 ATGFFDDVRVETADGQLLVIERPTIGSLNITAKMLQND-----IKNLESFGLAOSQY 123
QY 225 MD--LNGSIPRLRQ-----TALVAAVAVGYIDLSIIRNSIGEVDVYIIDL 269
|:|
Db 124 FNGATLNOAVAGLKEEYLGKGLNIQITPKYTKLAARRNRVDITIDCKSKAKITDIEFE- 182
QY 270 GEPYIDYRAVEVRGEGADKAFETVADEVPLLIGVFNHGKYEYKKNLJEMASAEHGYF 329
|:|
Db 183 GNOVYSDRKILMR-----QMSLTEGGIMTWLTFRSDRFDRQKFAQDMEKYVDFYQNNGYF 235
QY 330 DGRMLDSVDVILPDNADVSLIYDTGYRFDVYFETIDPKNOLTTDPDKIPVREL 389
|:|
Db 236 DFRLLDTIDQTNEDKTRQTKITVHEGGRFVKV---SIEGDNE-----VPKAE 283
QY 390 LEOLTYVMGEAVNLQAVRALSNDLIATRYFNMVNTEIVPEREIOINDQVSEFOSSSSR 449
|:|
Db 284 LEKLLTPMKPKWYERQGMTAVLG-----EIQNRKGSAGYATSS-- 320
QY 450 TEPAQVDESTLEPIVETVELTDGILMDISP-----IEFSASNLIDOKLNLVAAKARH 501
|:|
Db 321 -----EISVQPLPNAAGTQTVDFVLIIEPKRIYVNEIHITGNKTRDE--VYRRELQ 371
QY 502 LYDMPD-----RYLAINHDDGVNRSILGRISDAVSAVARAILPDESENEVIDLPE 552
|:|
Db 372 MESAPYDTSKLRQSKKEVELLGFEDNVQ-----FDVAVPAGTPDK-----VDL- 414
QY 553 RTALANRKPADVYQSKKVPYVAVASDKPRDGOIGL--GMGSDTGRLTKFEHNLINR 610
|:|
Db 415 NMSLTERST-----GSLDISAGWVDTG--LYMSAGVSODNL 449
QY 611 DGQAQAEIRLSDKKGVKLYATKPLSHPLNDQLRATLGYQGEVFGHSTNGFDLSTRTLE 670
|:|
Db 450 FCTGKSAALRASRSK--TTLNGSLSTFDPYFTADGVSLGY--DIYGAPFPRAASTSVKQ 505
QY 671 HEISRSIIONGGMNRTYSLRYRLDKLTKQAPPE-----TWQDLP---VDFVN--GK-PSQ 719
|:|
Db 506 YKTTTA---GGGVRMGIPVETEDRVNFGIAAEHLTVNVTYKAPKRYADFIKRYGTGDA 561
|:|
QY 720 EALLAGVAHVKTV-----ADNLVNPMPRGYRQRYSLSEVSSGLVSDANMAIARAGISGVY 773
```

```
Db      562 DGSFGLLYKGVGWRKTKDASWPTRGY-----LGVNAEIALPSSKIQY 608
QY      774 SEGDNAVGSNRAHQMTG-----GIOAGYIMSDNHNHYRLRFPAGDQ 817
Db      609 -----YSATHNQTWTFPLSKFTTLMLGGEVGIAGY---GRKKEIPFEFNFGGGLG 657
QY      818 SIRGAHDSLSPISSKGYLTGGQVLAAG---TAEYNYEFM-----KDLRLAVFGDI 865
Db      658 SVRGYESGTLGP---KYDEVGEKISYSGNKKANYSABELLFPMPGAKDARIVRLSLFADA 714
QY      866 G-----NAYDKGTNDTKIGAGVGRMASPVGOVAVDVAATGYK 903
Db      715 GSWMDGRITTAENGNKNSVYSENHAKSTFTNELRYSAGAVTWLSPUGPMKFTAYPLK 774
QY      904 E-EGNPIKLHFFIGTTPF 919
Db      775 KKPEDEIQRFQFQGLGTTTF 792
```

RESULT 6

AAB84745
ID AAB84745 standard; Protein; 792 AA.

AC AAB84745;

DT 17-SEP-2001 (first entry)

DE Amino acid sequence of a Neisseria gonorrhoeae protein.

KW Serogroup B protein; outer membrane protein; Neisserial infection;
vaccine.

OS Neisseria gonorrhoeae.

XX Key Location/Qualifiers

FT Peptide 1..21 /note= "signal peptide"

FT Protein 22..792 /note= "mature protein"

XX WO200152885-A1.

XX 26-JUL-2001.

PF 17-JAN-2001; 2001MO-IB00166.

PR 17-JAN-2000; 2000GB-0001067.

XX 09-MAR-2000; 2000GB-0005699.

PA (CHIR-) CHIRON SPA.

PI Pizsa M, Rappuoli R, Giuliani M;

DR WPI; 2001-451895/48.

XX N-PSDB; AAH42129.

PT Composition for treating or preventing infection to, detecting, or for
PT raising antibodies against Neisserial bacteria, comprises an N.
PT meningitidis serogroup B outer membrane preparation and an immunogenic
PT component -

XX Disclosure; Page 65-67; 83pp; English.

XX The present sequence represents a Neisseria gonorrhoeae protein. The

CC protein is used to produce the compositions of the invention. The

CC specification describes a composition, comprising a Neisseria

CC meningitidis serogroup B outer membrane preparation and an immunogenic

CC component. The immunogenic component is protein disclosed in WO99/57280,

CC WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/229412, WO95/03413,

CC WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA,

CC TbpA, TbpB, PilC, Opa, or Omp85. The composition is used for making

CC a medicament for treating or preventing infection due to Neisserial

CC bacteria; a diagnostic reagent for detecting the presence of Neisserial
CC bacteria or of antibodies raised against Neisserial bacteria; and/or
CC a reagent which can raise antibodies against Neisserial bacteria. It may
CC also be used as a vaccine.

XX Sequence 792 AA:

Query Match 5.6%; Score 265.5; DB 22; Length 792;

Best Local Similarity 20.9%; Pred. No. 1.1e-09;

Matches 192; Conservative 119; Mismatches 352; Indels 255; Gaps 39;

```
QY      116 TPLSLLEFLAQSSTENGTPND-----YIPYQSGPQSEVVPPTLPPEKGLIKRY 169
Db      16 SPLAADFPTIIDIIRVEGLQRTPESTVFNYLPVKYGGTYN-----DTHGSAIIKSYL 66
QY      170 ARLFNDGVNKPV---RLKAKFYSSSGSGETSAIGSSHOKTEPYANIKALED---ITQESA 224
Db      67 ATGFDDVAVETADGOLLITVIERPTIGSLNITGAKMLONDA---TKNLESFGLAQSY 123
QY      225 MD---LNGSIPRLRQ-----TALVARAVGYIDLSIRNSIGEVYIIDL 269
Db      124 FNOATLNOAVAGLKEEYILGRKLNIOITPKYTKLARNRVIDITIDEKSAKITDIEFE- 182
QY      270 GEPYVIDYRAVEVRBEGADKAFITVADEVPLIGVFHNGKYEKKNLLENASAHEGYF 329
Db      183 GNOVYSDRKLMR-----QMSLTREGIWTWLTFRSDREBQKFAODMEKVTDFYQNNGYF 235
QY      330 DGRWLDRSYDVILPNTADVSLIYDTGQYRFPEVEVFTIDKTNQTLTDPDLKPKREL 389
Db      236 DEFRIIDTQITNEDKTRQIKITVHEGGFRMGKV---SIEBDTNE-----VPKAE 283
QY      390 LEQLITVMKGEYNIQAVRALSNDLIATRYFMVNTEIYFPEREIOINDOVFEQSSSR 449
Db      284 LEKLLTMKRGKWEKQOMTAVLG-----ETONMGSGAGYAYS-- 320
QY      450 TPAQVDSSTLEPEVTEVELTDTGILMDISP-----IEFSASNLIODKMLVAKAKAH 501
Db      321 -----EISVQPLPNAGIKTYDFVLIHEPGKITYVNEHITGNKKTDE--VIRELRQ 371
QY      502 LYDMPDD-----RYLAINHDDGVNRSILGRISDAVAVARAILPDESENEYIDLE 552
Db      372 MESAPYDTSKLDQSKEREVELLGYFDNVQ-----FDAVPLAGTDPK-----VDL- 414
QY      553 RTALANKRTPADVYOSKVPVLYVFAASDKPRDQIGL--GMSDGTGTRLYTFEHNILNR 610
Db      415 NMSLTERST-----GSLDLASAGWQDTG--LYMSAGVSODNL 449
QY      611 DGYQAGAEELRLSEDKKGVKLYATKPLSHPLNDQLRATIGYQOEFGHSTNGFDLSTRILE 670
Db      450 FGTGKSALRLASRSK--TTNGSLSTFTDPTFTADGVSLGY--DIYGKAFDPKASTSVKQ 505
QY      671 HEISRSITQNGGMNFTYSRLRYRLDKLKTQAPPE---TWQDLR--VDENV--GK-PSQ 719
Db      506 YKTTTA---GGGVAMGIPVTEYDRVNGLAELHTLVNTYNNAPRYADFIKRYKGTGCA 561
QY      720 EALLAGVAVHKTU-----ADNLVPMRGYRGORYSLEVGSSGLVSDAMATARAAGISGY 773
Db      562 DGSFGLLYKGVGWRKTKDASWPTRGY-----LGVNAEIALPSSKIQY 608
QY      774 SEGDNAVGSNRAHQMTG-----GIOAGYIMSDNHNHYRLRFPAGDQ 817
Db      609 -----YSATHNQTWTFPLSKFTTLMLGGEVGIAGY---GRKKEIPFEFNFGGGLG 657
QY      818 SIRGAHDSLSPISSKGYLTGGQVLAAG---TAEYNYEFM-----KDLRLAVFGDI 865
Db      658 SVRGYESGTLGP---KYDEVGEKISYSGNKKANYSABELLFPMPGAKDARIVRLSLFADA 714
QY      866 G-----NAYDKGTNDTKIGAGVGRMASPVGOVAVDVAATGYK 903
Db      715 GSWMDGRITTAENGNKNSVYSENHAKSTFTNELRYSAGAVTWLSPUGPMKFTAYPLK 774
QY      904 E-EGNPIKLHFFIGTTPF 919
```


QY	321	NASASGHYEDGKWDLRSDVILPDPFADVSLIYDTGQYRDEDFEFTTIDKTNQGLTDP	360
	227	DEYQNNKGFIDRILDTIDTQIODEKTKQTIKITVEGGFFRGKV---SIGDITNE-----	278
QY	361	DKLPRKRELELDLTLVNNGEAYNLQAVFALSNDLILATRYFMNVTETIYPEREIOINDQY	440
DB	279	---VPKAELEKLLIMKRGKMYEQQMAYLG-----EIQNRMG	313
QY	441	SFEQSSSRTERPAQVDESTLEPIYETVELDGLIMDISP-----IEFSASNLIDKLT	492
DB	314	SAGVAYS-----EISVQPLPNAETKTYDVFVLHIEPGKAIYVNEIHITGNKKTDE-	363
QY	493	NLYAKAKHLIDMPD-----RVLIATHDQGVNRSLIGRISDAVSAVARAILPDES	543
DB	364	-VVRRELFQMESAPRDIOTSKLORSERVELLGYFTNVQ-----FDVAPLAGPDK-	411
QY	544	ENEVIDPERTALANRKTADVYOSKKVPLVYFVASDKPRDQIGL--GAGSPGTGRLVT	601
DB	412	---VDL--NMSLTERST-----GSIDLDSAGVQDTG--LVN	440
QY	602	KFEHNLINRDGYQAGAEILRLSEDKKGVKLYATKPLSHPLNDOLRATLGYQOEFGHSTNG	661
DB	441	SAGVYQDMVFEGKSAALRAASRK--TTLNGLSLFTDPTPYFADGVSLGY--DYVYKAFDP	496
QY	662	FDLSIRLIEHETSRLIIONSGMNR-----YLSLRYRLDKLKTQAPRPTNQDLP---VDVY	713
DB	497	RKASTISIQYTTTR---GAGIRMSVPYETDYRVNFGVLAENHLYVTYTKAKPRHYADFT	552
QY	714	N--GK-PSQEALLAGVAHVKTV-----ADNLVNPMRGYPORYSLSEVGSSGLVSDANMAI	764
DB	553	KYKGTDTGDTGSPFGMWLTKGVYGNKRNKTDALMPTGY-----LTGVAAEI	599
QY	765	ARAGISGYVSYSGDNAYGNSRAHQMTG-----GIQAGYIMSDNFNNHVPYR	808
DB	600	ALPGSKLOY-----YSATHNOTWFEPLSKPTTLMIGEVEGLAGSY--GRTKEIPFF	648
QY	809	LRFPAQGQOSIRGVAHDSILSPDKGYLTGQVLAAG--TAENYEFM-----KD	856
DB	649	ENFYGGGIGSIRGIESGTLP--KVYDEYGEKISTGYGNKKANVASALLEPMPGAKDART	705
QY	857	LRLAVFGDIGNAYD-----KGFNTDKIGAGVGYRMAS	889
DB	706	VRLSLFADAGSVMPGKTYDDNSSSATGGRVQNYIGAGNTHTKSTFTYNLRLRSAGAGATWLS	765
QY	890	PVGQVRDVATGAKRE--EGNPFIKLHFTIGPFF	919
DB	766	PLGPKKFSYAPRLKKKPEDEIQRFOFOLGPTTF	797
RESULT 12			
AAB84744			
ID	AAB84744 standard; protein: 797 AA.		
XX	AAB84744;		
XX	17-SEP-2001 (first entry)		
DE	Amino acid sequence of a Neisseria serogroup B protein.		
XX	Serogroup B protein; outer membrane protein; Neisserial infection;		
KW	vaccine.		
XX	Neisseria meningitidis.		
OS			
XX	Key		
FT	Peptide		
FT	Protein		
XX	Location/Qualifiers		
XX	1..21		
FT	/note="signal peptide"		
FT	22..797		
XX	/note="mature protein"		
XX	W0200152885-A1.		

PD 26-JUL-2001.
XX
XX
XX 17-JAN-2001; 2001MO-IB00166.
XX
XX 17-JAN-2000; 2000GB-0001067.
XX 09-MAR-2000; 2000GB-0005699.
XX
XX (CHIR-) CHIRON SPA.
XX
XX
XX Pizsa M, Rappuoli R, Giuliani M;
XX
XX WPI: 2001-451895/48.
XX
XX N-PSDB; AAH42128.
XX
XX
XX Composition for treating or preventing infection to, detecting, or for
XX raising antibodies against Neisserial bacteria, comprises an N.
XX meningitidis serogroup B outer membrane preparation and an immunogenic
XX component -
XX
XX Disclosure, Page 59-61; 83pp; English.

CC The present sequence represents a *Neisseria* serogroup B protein. The
CC protein is used to produce the compositions of the invention. The
CC specification describes a composition, comprising a *Neisseria*
CC meningitidis serogroup B outer membrane preparation and an immunogenic
CC component. The immunogenic component is protein disclosed in WO99/57280
CC WO99/36544, WO99/42578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
CC WO99/31132, WO99/58683, WO99/55873, and/or *N. meningitidis* protein PORa
CC TbpA, TbpB, PilC, Opa, or OmpB5. The composition is used for making
CC a medicament for treating or preventing infection due to *Neisserial*
CC bacteria; a diagnostic reagent for detecting the presence of *Neisserial*
CC bacteria or of antibodies raised against *Neisserial* bacteria; and/or
CC a reagent which can raise antibodies against *Neisserial* bacteria. It may
CC also be used as a vaccine.

SQ Sequence 797 AA;

Query Match	5.5%;	Score 259;	DB 22;	Length 797;
Best Local Similarity	20.5%;	Pred. No. 3.2e-09;		
Matches 191;	Conservative 121;	Mismatches 342;	Indels 278;	Gaps 40;

```

QY 116 TP$LSLEELFAQES$TEMG$INPMD-----Y$PEQ$EQ$PN$SEVVP$P$LL$EP$K$ELL$RL$Y 16
Db 16 $P$LA$D$T$D$D$IR$VE$G$Q$RT$E$P$T$VE$N$Y$LL$P$V$K$V$D$H$Y$N$-----D$H$G$A$T$IK$SL$Y 66
QY 170 AR$F$ND$Q$V$N$K$V$P$-----RL$K$A$F$T$Q$S$O$G$E$N$S$A$IC$S$H$Q$K$E$P$Y$A$N$IK$A$A$E$D$--IT$Q$E$S$A 224
Db 67 AT$F$E$D$D$V$R$E$T$A$D$Q$G$Q$LL$T$V$T$E$R$P$T$IG$SL$N$IT$G$AK$M$Q$N$D$A$-----IK$KN$E$S$E$P$L$A$Q$S$Q$Y 128
QY 225 MD$-----L$N$G$S$T$R$L$F$Q$-----T$A$Y$A$N$A$V$G$Y$D$D$LS$T$IN$N$S$IG$E$V$Y$T$IH$D$ 266
Db 124 F$N$Q$AT$D$M$O$A$V$G$L$E$E$Y$L$G$R$G$K$N$L$IQ$IT$P$K$T$A$R$N$R$H$D$ID$IT$D$G$K$S$AK$IT$D$IE$F$E- 1828
QY 270 G$E$P$Y$ID$R$A$E$V$G$E$G$AD$K$A$F$T$V$A$D$E$V$LL$G$D$V$-----F$H$G$A$E$T$K$N$L$IE 320
Db 183 G$N$O$Y$S$D$R$K$LM$R$-----Q$M$SL$T$G$G$I$W$T$ML$T$R$N$O$F$N$E$O$K$F$A$D$M$E$K$V$T 226
QY 321 N$A$S$A$H$G$Y$F$D$G$R$M$LD$R$S$V$D$V$LL$P$N$T$A$D$V$SL$Y$D$T$G$Y$R$E$D$E$V$F$T$D$P$K$T$Q$N$Q$LL$T$D$P 360
Db 227 D$F$Y$O$N$N$G$E$F$D$R$ID$F$D$T$Q$T$N$E$D$K$T$Q$IT$K$IT$E$G$E$G$R$F$R$G$K$V$--S$IE$G$D$T$N$E$----- 278
QY 381 D$K$L$V$K$R$E$LL$E$O$LL$T$V$M$G$E$N$Y$N$Q$A$R$A$LS$N$D$L$A$R$Y$N$M$V$T$E$L$V$P$E$R$Q$I$O$N$D$Y 440
Db 279 -----V$P$A$E$LE$K$LL$T$M$P$G$K$Y$E$R$Q$O$T$A$Y$G$-----E$IO$N$R$M$G 313
QY 441 S$F$E$O$S$S$R$T$E$P$A$Q$V$D$E$S$T$E$P$Y$E$T$V$E$L$D$G$IL$M$D$P$-----I$E$F$S$A$N$L$I$O$D$K$L 4922
Db 314 S$A$G$Y$A$Y$S$-----E$IS$Q$P$L$P$N$A$E$T$K$V$D$P$V$L$H$E$P$R$K$T$I$Y$V$N$E$IH$IT$G$N$K$R$T$D$E- 363
QY 493 N$V$A$A$K$R$H$L$Y$D$M$P$D$-----R$V$L$A$N$H$D$D$Q$V$N$R$S$T$L$G$R$IS$D$A$V$A$V$A$R$A$LL$P$D$E$S 543
Db 364 -V$V$R$E$T$R$O$M$E$S$A$Y$D$T$S$K$O$R$K$E$R$E$LL$G$Y$D$N$O$-----F$D$A$V$P$L$A$G$P$P$D$K- 411

```



```

DB      600 ALPGSKLQY-----YSATHNQTFPPFLSKFTTLMLGGEVGIAGY---GRTKEIPFF 648
OY      809 LRFEPAGDOSIRGYAHDLSPISDKGYLTGGQYLAVG---TAENYEFM-----KD 856
DB      649 ENFYGGGLGSVNGYESGTLGP---KYDEYGEKISYGNKKANVASALLEPMPGADART 705
OY      857 LRLAVFGDIGNAYD-----KGFTNDTKIGAGVGVWAS 889
DB      706 VRLSLFADAGSVWDGKTYDDNSSATGGRVQNTYGAGNTHKSTFTNELRYSAGCAVTWLS 765
OY      890 PYGQVRVDVATGVKE---EGNPILKHFIFIGTF 919
DB      766 PLGPMKFSYAYPLKKRPEDEIQRFQGLGTTTF 797

RESULT 14
AAB23784
ID      AAB23784 standard; Protein; 797 AA.
XX
AC      AAB23784;
XX
DT      12-JAN-2001 (first entry)
XX
DE      Neisseria meningitidis serogroup B amino acid sequence.
XX
KW      CG motif; complete Freund's adjuvant; phosphorothioate; immunogenic;
KW      Neisseria antigen; Neisseria meningitidis; Neisseria gonorrhoeae;
KW      bactericidal; antibacterial; vaccine; immunostimulatory; infection;
KW      immune response.
XX
OS      Neisseria meningitidis.
XX
PN      WO20050075-A2.
XX
PD      31-AUG-2000.
XX
PF      09-FEB-2000; 2000WO-IB00176.
XX
PR      26-FEB-1999; 99US-0121792.
XX
PA      (CHIR-) CHIRON SPA.
XX
PI      Grandi G, Rappuoli R, Giuliani MM, Pizsa M;
XX
DR      WPI; 2001-015529/02.
XX
PT      Immunogenic composition useful for stimulating an immune response in a
PT      mammal against Neisseria infection, comprises Neisseria antigen and an
PT      adjuvant composition comprising an oligonucleotide with a CG motif -
XX
PS      Claim 22; Page 32; 39pp; English.
XX
CC      The present invention describes an immunogenic composition (I)
CC      comprising a Neisseria antigen and an adjuvant composition comprising an
CC      oligonucleotide comprising at least 1 CG motif. Also described is an
CC      adjuvant composition (II) comprising an oligonucleotide which comprises
CC      at least 1 CG motif and a complete Freund's adjuvant (CFA), where the
CC      oligonucleotide preferably comprises at least one phosphorothioate bond.
CC      AA92359 to AA92385 represent specifically claimed oligonucleotides of
CC      the present invention. (I) is useful for stimulating an immune response
CC      in a mammal, preferably a human, against Neisseria infection, preferably
CC      Neisseria meningitidis infection, and in the manufacture of a medicament
CC      for inducing a protective immune response in a mammal. The present
CC      sequence represents the claimed Neisseria meningitidis serogroup B amino
CC      acid sequence disclosed in GB-9928197.4, which is given in the present
CC      invention.
XX
SQ      Sequence 797 AA;

Query Match 5.5%; Score 259; DB 22; Length 797;
Best Local Similarity 20.5%; Pred. No. 3.2e-09;
Matches 191; Conservative 121; Mismatches 342; Indels 278; Gaps 40;
```

```

OY      116 TPLSLEELFAOESTEMGINFND-----YIPXQGEOPNEVVPPTLEPERKPELLIKRLY 169
DB      16 SPLALADFTIODIRVEGLQRTPESTVENYLPVKGDITYN-----DTHGSAIIRKSLY 66
OY      170 ARLEFNDGVNKP---RLKAKFYOSSGSGETSAIGSSHOKTEPPYANITAALED--ITDESA 224
DB      67 ATGFEDDVARETADGQLLTVIERPTIGSLNITGAKMLONDA---IKRNLSEFSLAOSQY 123
OY      225 MD--LNGSIPRLRO-----TALYAARAVGYVIDLSIINNSIGEYVYIIDL 269
DB      124 FNOATLQAVAGLKEEYLGKGLNIQTTPYTKLARNRVIDITIIDCKSAKITDIEFE- 182
OY      270 GEPYIDYRAVEVREGADDAKFTVADEVPPLIGDV-----PHHGKETRKNLIE 320
DB      183 GNOYVSDRKLNR-----QMSLTFEGGIWTMLTRSNQNEQKFAQDMKEVT 226
OY      321 NASAEHGYFDGRWLDRSDVYLIPDNTADVSLIDTGYQYRFDEVYFTTIDPKTNQLTDP 380
DB      227 DFYQNNGYFDFRILDTDIDQNEDEKTKOTIKITVHEGGRFRGKV---SIEGDITNE----- 278
OY      381 DKLPYKRELLEOLITVMMGEAYNLOAVRALSNDLIATRFEMVMTETIVPEREQIONDOV 440
DB      279 ---VPRAELEKLLTMKPGKWKYERQOMTAVLG-----EIQNRMG 313
OY      441 SFEQSSSRTEPAQVDESTEPEVIEVELTDGILMDISP-----IEFSASNLIODKL 492
DB      314 SAGAVYS-----EISQVPLNAETKTVDFVLHIEPGKIIYVNEIHITGNKKTDE- 363
OY      493 NLVAAKARHLYDMPD-----RVLAINHDDGVNRSILGRISDAVSAVARAILPDES 543
DB      364 -VVRRELQWESAPYDTSKLQRSKERVELLGFEDNVQ-----FDAVPLAGTFDPK- 411
OY      544 ENEVYIDPERTALANRKTPTADVQSKVPLXVYVASQKPRDQIGL--GMSDPTGTLVY 601
DB      412 ---VDL--NMSLTERST-----GSIDLSAGWYQDTG--LVM 440
OY      602 KFEHNLINRDYQAGALRLSEDKKGKLYATKPLSHPLNDQLRATGYOQOEVGSHSTG 661
DB      441 SAGVSDQNLFGTGKSALRLASRSK--TTLNGSLSTPTPTFADQVSLGY--DYVGKAFDP 496
OY      662 FDLSTRTLEHISRSIITQNGGWR---TYSLRYLRLDKLKTQAPPTWDLR---VDVY 713
DB      497 RKASTSIQYKTTTA---GAGIRMSVPTVEYDRVNGLVAEHLTVTNKAPKHVADFI 552
OY      714 N--GK-PSQALLAGVAVHKTV-----ADNLVPMMGYQORSLEBGSGLVSDAMMAI 764
DB      553 KKYGTIDGTGSRKGLWKYGTGGRNKTDSALMPTRGY-----LTGVAEI 599
OY      765 ARAGISGVYSGDNAGSNRAHQMTG-----GIQAGYIWSDNFNHVPYR 808
DB      600 ALPGSKLQY-----YSATHNQTFPPFLSKFTTLMLGGEVGIAGY---GRTKEIPFF 648
OY      809 LRFEPAGDOSIRGYAHDLSPISDKGYLTGGQYLAVG---TAENYEFM-----KD 856
DB      649 ENFYGGGLGSVNGYESGTLGP---KYDEYGEKISYGNKKANVASALLEPMPGADART 705
OY      857 LRLAVFGDIGNAYD-----KGFTNDTKIGAGVGVWAS 889
DB      706 VRLSLFADAGSVWDGKTYDDNSSATGGRVQNTYGAGNTHKSTFTNELRYSAGCAVTWLS 765
OY      890 PYGQVRVDVATGVKE---EGNPILKHFIFIGTF 919
DB      766 PLGPMKFSYAYPLKKRPEDEIQRFQGLGTTTF 797

RESULT 15
AAB23788
ID      AAB23788 standard; Protein; 797 AA.
XX
AC      AAB23788;
XX
DT      12-JAN-2001 (first entry)
```

[illegible]

[illegible]

```
XX Porphyromonas gingivalis; PG: periodontal disease; gingivitis;
KM vaccine; antigenic.
XX
XX Porphyromonas gingivalis.
XX
XX MO929870-A1.
XX
XX 17-JUN-1999.
XX
XX 10-DEC-1998: 98MO-AU01023.
XX
XX 04-AUG-1998: 98AU-0005028.
XX 10-DEC-1997: 97AU-0000839.
XX 31-DEC-1997: 97AU-0001182.
XX 30-JAN-1998: 98AU-0001546.
XX 10-MAR-1998: 98AU-0002264.
XX 09-APR-1998: 98AU-0002911.
XX 23-APR-1998: 98AU-0003128.
XX 05-MAY-1998: 98AU-0003338.
XX 22-MAY-1998: 98AU-0003654.
XX 29-JUL-1998: 98AU-0004917.
XX
XX (CSLC-) CSL LTD.
XX
XX Agius CT, Barr IG, Hocking DM, Margetis MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
XX MPI: 1999-385613/32.
XX N-PSDB; AAX91724.
XX
XX Antigenic Porphyromonas gingivalis peptides for preventing
XX gingivitis
XX
XX Claim 1: Page 487-489: 588pp: English.
XX
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX AAY34583. AAX91802 to AAX91889 represent PCR primers used in the
XX isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX activity with a vaccine mechanism of action. The PG polypeptides can be
XX used as vaccines especially against Porphyromonas gingivalis. Probes can
XX be used to detect Porphyromonas gingivalis in standard hybridisation
XX assays. Porphyromonas gingivalis is involved in periodontal disease
XX especially gingivitis.
XX
XX Sequence 775 AA:
SQ
Query Match 4.7%; Score 224.5; DB 20; Length 775;
Best Local Similarity 20.6%; Pred. No. 7.4e-07;
Matches 155; Conservative 90; Mismatches 290; Indels 217; Gaps 34;
QY 300 PLLIDGVFNHGKYEKKNLLENASAEHGFDDGWLDRSDVILPDN-TADVSLIYDTGTQ 358
DB 109 PFVLSOV-----SDSRAKATNLRHGHYFDAK-VKSVYTTLKDKSLAKISYVDWASP 163
QY 359 YRFDEVEFTIDPKTNQLTDPDKLPVKRELLQGLTVMNGEAYNL-----QAVRALS 411
DB 164 YHVDSTIRPFI-----STFSDSIATYKOTPS---LIRKGDQFNLAKEHERQTISALL 213
QY 412 NDLIATRFNMVNTLIVFERBOIQDOVSFEQSSSSRTPEA-----QVDESTL----- 460
DB 214 RD---NGYFFRPDRIYADPLLVGAVCLRAKLSSEDTFRPQAMRPWRIGKRAVLGGMN 270
QY 461 -EPVIEIVELTD-GIL---MDISP-----IEFSASNLIDQKLNIVAAKARHLXDPDD 508
DB 271 GSPSDSLEVEEMKALYYKRMFVRPIIAKRRFFSGNLYRQK----- 313
QY 509 RVLAINHDDGVRSILIGRISDAVSAVARAILPDESENEVIDLPERTAL-----ANKTPA 563
DB 314 -----DDETRKSLARIG-AFSYIDLNFLORDSISGLLDVRLTLTTLDKPMDSLET-- 363
QY 564 DVYOSKK-----VPLYVFVASDKPRDGOIGLGM---GS---DTGTRLVYKFEHNLINRDG 612
```

```
DB 364 -LFTSKSNDFIGPGLNFALARNVFGGGENLSWMNIGSSYEMETGNR-PENSSNRLIDINS 421
QY 613 YQAGAEIRLSEDKKGVKLYATKPLSHPIINDQLRAT----- 647
DB 422 YNMNTAVNLSPFSIVFPGLDKYRYYPPTTTFQASATALNBAHFYSMSYFSGFTTYEPQ 481
QY 648 -----LGXQDEVGHSTNGFDLSTRT---LEHISRSIIIONGMNRTISLRVLD 694
DB 482 SKEHRHAIPLKLVNLLGHOTETFOAITANNPPLLSLOQFALQOMGYITTFN-----K 536
QY 695 KLKTQAPPETWQDLPVDFVNGKPSQAEALLAGVAHKYVADLVN-PMRGY-----ROBS 748
DB 537 SVSEKSPHLLMQFGLSAGNLLNLIYLAACKKXSDT--KRFVGVPSQFKAIGELRYS 594
QY 749 LEVSSSGIVSDANAAIA-RAGISGVSEGDNAYSNRAHMTGTGIQAGYIWSDFNHVY 807
DB 595 YTI-----DRNOSLATRFGTGVYISYG-----NMRAVAPY 623
QY 808 RLRFAGGDOSIRGYAHDSL-----SPISDKGYL---TGQVLAVGTRAEYFEKMDLRL 859
DB 624 SEQFYVGANSIRAFYVRSIGRFPNPSDMQSYLDQVGEFKLEANVEYRGKLFEGDLHA 683
QY 860 AVFEGDIGNAY-----DKGFTNDTRIGAGVGRMASPVQGVFVDVATG 901
DB 684 AVFDAGNWNLLRSDSSRPGALSEVGSYSNFIINSIALGTGVGLRDLAFLVRAVDVFG 743
QY 902 VK-----EEG-----NPIKLFPIGTGPF 919
DB 744 LHLPEYNTGKKGYNIIPRFKDAIGFHLAVGYPF 775
RESULT 20
AAY34379
ID AAY34379 standard; Protein: 803 AA.
XX
XX AAY34379:
XX
AC 25-AUG-1999 (first entry)
XX
DE Porphyromonas gingivalis protein PG45.
XX
KM Porphyromonas gingivalis; PG: periodontal disease; gingivitis;
XX vaccine; antigenic.
XX
OS Porphyromonas gingivalis.
XX
XX MO929870-A1.
XX
XX 17-JUN-1999.
XX
XX 10-DEC-1998: 98MO-AU01023.
XX
XX 04-AUG-1998: 98AU-0005028.
XX 10-DEC-1997: 97AU-0000839.
XX 31-DEC-1997: 97AU-0001182.
XX 30-JAN-1998: 98AU-0001546.
XX 10-MAR-1998: 98AU-0002264.
XX 09-APR-1998: 98AU-0002911.
XX 23-APR-1998: 98AU-0003128.
XX 05-MAY-1998: 98AU-0003338.
XX 22-MAY-1998: 98AU-0003654.
XX 29-JUL-1998: 98AU-0004917.
XX
XX (CSLC-) CSL LTD.
XX
XX Agius CT, Barr IG, Hocking DM, Margetis MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
XX MPI: 1999-385613/32.
XX N-PSDB; AAX91597.
XX
XX Antigenic Porphyromonas gingivalis peptides for preventing
```



```
OY 559 RKTPA-----DYQSK-----VPLYVFASDKP 582
      : : : : :
Db 168 WQPVGSDFDODSWENKSTVLGAVTRKAYPLAKGNTRAVNDPTATLVNVDSCR 227
OY 583 RDGQIGLGWSDTGR-----LYTKFEHNLINRDGYAGAEI- 619
      : : : : :
Db 228 ----IAFGDFEITGTQRYPEQIVSGLARFQPGTPYDLDLDFQOAL-EQNGHYSGASVQ 282
      : : : : :
OY 620 ----RLSEDKKGYLTKPLSHPLNDQLRAT-----LG-----YQOEVEGHSTNG 661
      : : : : :
Db 283 ADFDRLOGDVPRVAVSTEVKRNKLETGIRLDSYGLGKRAYDYNNLFNKGYIGSVYMD 342
OY 662 FDLSTRTLEHRSISIIQNGW---NRTYSLRRLDKLKTQAPPETMODLPVDFVNGKPS 718
      : : : : :
Db 343 MDRYETTLAAGISQPRNRYGNWTSNVSYNSTQNLKRAFSGGCIWVRDRAIDARLG 402
OY 719 QEALLAGVAV-----HKT-----VADNLVNPFRGYRQYSLVSGSSGLVSD 759
      : : : : :
Db 403 AEFLAEGRKIPGSDIDIGNSHATMLTASMKRQLNNVLHPENGHYLDOKIGTTLGAFIJS 462
OY 760 ANN-ALARAGISGVYSGDNAYGSNRAHOMTGTI---QAGY-IMSDNENHVPYRLRF 812
      : : : : :
Db 463 TALIRTSARAG-----YFTPEKKLGTFTIRGQAGTVARDNAN-VPSGLMFR 510
OY 813 AGGDQISIRGVAHDSLSPISDKGYLTGGQVLAAGTAENYEFMKDLRLAVFGDIGNAYDKG 872
      : : : : :
Db 511 SGGASSVRYGELDISIGLAPNGSVLPERALLVGSLEQOLPPTRTLSCGAVFHDMDGDAANF 570
OY 873 FTMDDTIGAGVGRMASPVGOVRVDVATGVKEEGNPILKHFFIGTPF 919
      : : : : :
Db 571 KRMKLHSGLSGLVGRWFSPLAFPSFDIAYGHSDK--KIRMHISLGTFR 615
```

RESULT 22

AA99623
ID AAY99623 standard; Protein; 609 AA.

XX AAY99623:

XX 03-JAN-2001 (first entry)

DE Neisseria meningitidis BASB040 putative protein sequence #1.

KW BASB040; bacterial disease; respiratory tract infection; bacteraemia;
meningitis; cancer; autoimmune disease.

OS Neisseria meningitidis.

PN WO200034480-A1.

PD 15-JUN-2000.

PF 02-DEC-1999; 99WO-EP09560.

PR 07-DEC-1998; 98GB-0026886.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Ruelle J;

XX WPI: 2000-423426/36.

DR N-PSDB; AAA48507.

PT Novel BASB040 polypeptides of Neisseria meningitidis useful for
diagnostic, prophylactic and therapeutic purposes against microbial
diseases comprise a specific amino acid sequence

XX Claim 4; Page 59-60; 98pp; English.

CC The present sequence is a putative version of the Neisseria
meningitidis strain ATCC13090 BASB040 protein sequence. This protein is
similar to the D15 outer membrane protein of the bacterium. The protein.

CC its gene, antibodies, antagonists and agonists can be used to diagnose
CC and treat bacterial diseases such as those leading to upper respiratory
CC tract infections, bacteraemia and meningitis. In addition, they can be
CC used in vaccines for use against cancer and autoimmune diseases.

SO Sequence 609 AA;

Query Match 4.6%; Score 219; DB 21; Length 609;
Best local similarity 22.0%; Pred. No. 1.3e-06;

Matches 130; Conservative 74; Mismatches 232; Indels 154; Gaps 18;

```
OY 462 PVIEVELTDGILMDISPIEFSSASNLIQDKNLVAAKRHLVYMDPDRVLAINHDDGYNR 521
      : : : : :
Db 42 PDTESVRLKPRFPRIDPQDSEIKDWEHLPLITQOQEEVLDRQGTFLAEAPDNVKT 101
OY 522 SI--LGRISDAVSA-----VARAILPDESENVIDLPETALAN 558
      : : : : :
Db 102 MLRSKGYFSSKVSLETKDGAVTHTPGPRTKIANVGAIIIGDILSDGNIAEYRNALEN 161
OY 559 RKTPA-----DYQSK-----VPLYVFASDKP 582
      : : : : :
Db 162 WQPVGSDFDODSWENKSTVLGAVTRKAYPLAKGNTRAVNDPTATLVNVDSCR 221
OY 583 RDGQIGLGWSDTGR-----LYTKFEHNLINRDGYAGAEI- 619
      : : : : :
Db 222 ----IAFGDFEITGTQRYPEQIVSGLARFQPGTPYDLDLDFQOAL-EQNGHYSGASVQ 276
OY 620 ----RLSEDKKGYLTKPLSHPLNDQLRAT-----LG-----YQOEVEGHSTNG 661
      : : : : :
Db 277 ADFDRLOGDVPRVAVSTEVKRNKLETGIRLDSYGLGKRAYDYNNLFNKGYIGSVYMD 336
OY 662 FDLSTRTLEHRSISIIQNGW---NRTYSLRRLDKLKTQAPPETMODLPVDFVNGKPS 718
      : : : : :
Db 337 MDRYETTLAAGISQPRNRYGNWTSNVSYNSTQNLKRAFSGGCIWVRDRAIDARLG 396
OY 719 QEALLAGVAV-----HKT-----VADNLVNPFRGY-----RQRYSLVSGS 753
      : : : : :
Db 397 AEFLAEGRKIPGSDIDIGNSHATMLTASMKRQLNNVLHPENGHYLDOKIGTTLGFIJS 456
OY 754 SGLVSDANMAIARAGISGVYSGDNAYGSNRAHOMTGTI---QAGYIMSDNENHVPYRL 809
      : : : : :
Db 457 TALIRTS---ARAG-----YFTPEKKLGTFTIRGQAGTVARDNADVPISGL 501
OY 810 RFFAGDQISIRGVAHDSLSPISDKGYLTGGQVLAAGTAENYEFMKDLRLAVFGDIGNAY 869
      : : : : :
Db 502 MFRSGGASSVRYGELDISIGLAPNGSVLPERALLVGSLEQOLPPTRTLSCGAVFHDMDGDA 561
OY 870 DKGFTNDTKIGAGVGRMASPVGOVRVDVATGVKEEGNPILKHFFIGTPF 919
      : : : : :
Db 562 ANFRMKLKHGSLGSLVGRWFSPLAFPSFDIAYGHSDK--KIRMHISLGTFR 609
```

RESULT 23

AA99624
ID AAY99624 standard; Protein; 609 AA.

XX AAY99624:

XX 03-JAN-2001 (first entry)

DE Neisseria meningitidis BASB040 putative protein sequence #2.

KW BASB040; bacterial disease; respiratory tract infection; bacteraemia;
meningitis; cancer; autoimmune disease.

OS Neisseria meningitidis.

PN WO200034480-A1.

PD 15-JUN-2000.

PF 02-DEC-1999; 99WO-EP09560.

XX

```

PR 07-DEC-1998: 98GB-0026886.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Ruelle J:
XX
XX WPI: 2000-423426/36.
XX N-PSDB: AAA48508.
XX
PT Novel BASB040 polypeptides of Neisseria meningitidis useful for
PT diagnostic, prophylactic and therapeutic purposes against microbial
PT diseases comprise a specific amino acid sequence
XX
XX Claim 4: Page 60-61; 98pp; English.
XX
XX The present sequence is a putative version of the Neisseria
XX meningitidis strain ATCC13090 BASB040 protein sequence. This protein is
XX similar to the D15 outer membrane protein of the bacterium. The protein,
XX its gene, antibodies, antagonists and agonists can be used to diagnose
XX and treat bacterial diseases such as those leading to upper respiratory
XX tract infections, bacteraemia and meningitis. In addition, they can be
XX used in vaccines for use against cancer and autoimmune diseases.
XX
XX Sequence 609 AA:
XX
Query Match 4.6%; Score 219; DB 21; Length 609;
Best Local Similarity 22.0%; Pred. No. 1.3e-06;
Matches 130; Conservative 74; Mismatches 232; Indels 154; Gaps 18;
OY 462 PVEYVELDGLIMDISPIEFSASNLIQDKLNVAAKARHLVMPDDRVLAINHDDGVR 521
DB 42 PDIESVYKLRKPFVRLIDTQDSEIKDMVEEHLPLITQOQEEVLDEKQGFLEAPDVKVT 101
OY 522 SI--LGRISDAVSA-----VARALLPSESENEVIDLPERALAN 558
DB 102 MLRSKGYFSSKYSLTEKDGAYVYHITPGPRTKIANGVALLGILSDGNIAEYRANLEN 161
OY 559 RKTPTA-----DYYOSKK-----VPLYVFVASDKP 502
DB 162 WOQPVGSDPDDQSWENSKYSVLGAVTRKAYPLAKLGNTRAAVNPDTATADLVNVVDSGRP 221
OY 583 RQOQIGLGMGSDTGR-----LVTKFEHNLINRDGYOAGAEI- 619
DB 222 ----IAFGDFEITGTQRYPEQIVSGIARFQPGMPYDLDDLDFQOAL-BQNGHYSGASVQ 276
OY 620 ----RLSEDKKGVKLYATKRLSHPLNDOLRAT----LG-----YQOEYFGHSTNG 276
DB 277 ADPDRLOGDRVVPKVSVEVKRHKLETGIRLDSEYGLGKIAVDYNNLKNKGYIGSVWD 336
OY 662 FDLSTRITLHEHSRSTIIONGWM---NRTYSLRYRLDKLKTQAAPPETWODLPVDFVNGKFS 718
DB 337 MOKYETTLAAGISQPRNRYNGNWTNSVSTNQTQNLERKRAFSGGIWMVDRAGIDARLG 396
OY 719 QEALLAGVAV-----HKT-----VADNLVPMRGY-----RQRYSLVSGS 753
DB 397 AEFLEGRKIPGSDIDLGNHATMLTASWKRLQLNNVLRHPNGHYDKIGITGLTFGLSS 456
OY 754 SGLVSDAMMATARAGISGVYSGFDNAYGSRNRAHQMTGGI----QAGYIWSDFNHNVPYRL 809
DB 457 TALLRTS---ARAG-----YFETPENKKLGTFEIIINGQAGYVARONADVPSSL 501
OY 810 RFPAAGDQSIKGYAHDSLSPISDKGYLGGQVLAAGTAEVNEFKMDRLAVFGDIGANV 869
DB 502 MFRSGASSVAGYEIYDLSGLAGPNSVLPERRALLVGSLEYQLEPTRLTJSGAIFHMDGMA 561
OY 870 DKGFTNDTKIAGVGVNMAVPGQVRVDVATGVKKEGNPIKLFHFIPTGF 919
DB 562 ANFKRMKLKHGSGLGVRWFSPLAPSFDAIYGHSDK--KIRHISLGITRF 609

```

```

XX
XX AC AAY99625;
XX
XX 03-JAN-2001 (first entry)
XX
XX Neisseria meningitidis BASB040 putative protein sequence.
XX
XX DE Neisseria meningitidis BASB040 putative protein sequence.
XX
XX KW BASB040; bacterial disease; respiratory tract infection; bacteraemia;
XX meningitis; cancer; autoimmune disease.
XX
XX OS Neisseria meningitidis.
XX
XX PN WO200034480-A1.
XX
XX PD 15-JUN-2000.
XX
XX PF 02-DEC-1999; 99WO-EP09560.
XX
XX PR 07-DEC-1998; 98GB-0026886.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX PI Ruelle J:
XX
XX WPI: 2000-423426/36.
XX N-PSDB: AAA48509.
XX
XX Novel BASB040 polypeptides of Neisseria meningitidis useful for
XX diagnostic, prophylactic and therapeutic purposes against microbial
XX diseases comprise a specific amino acid sequence
XX
XX Claim 4: Page 62; 98pp; English.
XX
XX The present sequence is a putative version of the Neisseria
XX meningitidis strain H44/76 BASB040 protein sequence. This protein is
XX similar to the D15 outer membrane protein of the bacterium. The protein,
XX its gene, antibodies, antagonists and agonists can be used to diagnose
XX and treat bacterial diseases such as those leading to upper respiratory
XX tract infections, bacteraemia and meningitis. In addition, they can be
XX used in vaccines for use against cancer and autoimmune diseases.
XX
XX Sequence 587 AA:
XX
Query Match 4.5%; Score 215; DB 21; Length 587;
Best Local Similarity 22.1%; Pred. No. 2.3e-06;
Matches 132; Conservative 73; Mismatches 225; Indels 166; Gaps 20;
OY 462 PVEYVELDGLIMDISPIEFSASNLIQDKLNVAAKARHLVMPDDRVLAINHDDGVR 521
DB 20 PDIESVYKLRKPFVRLIDTQDSEIKDMVEEHLPLITQOQEEVLDEKQGFLEAPDVKVT 79
OY 522 SI--LGRISDAVSA-----VARALLPSESENEVIDLPERALAN 558
DB 80 MLRSKGYFSSKYSLTEKDGAYVYHITPGPRTKIANGVALLGILSDGNIAEYRANLEN 139
OY 559 RKTPTA-----DYYOSKK-----VPLYVFVASDKP 582
DB 140 WOQPVGSDPDDQSWENSKYSVLGAVTRKAYPLAKLGNTRAAVNPDTATADLVNVVDSGRP 199
OY 583 RQOQIGLGMGSDTGR-----LVTKFEHNLINRDGYOAGAEI- 619
DB 200 ----IAFGDFEITGTQRYPEQIVSGIARFQPGMPYDLDDLDFQOAL-BQNGHYSGASVQ 254
OY 620 ----RLSEDKKGVKLYATKRLSHPLNDOLRAT----LG-----YQOEYFGHSTNG 661
DB 255 ADPDRLOGDRVVPKVSVEVKRHKLETGIRLDSEYGLGKIAVDYNNLKNKGYIGSVWD 314
OY 662 FDLSTRITLHEHSRSTIIONGWM---NRTYSLRYRLDKLKTQAAPPETWODLPVDFVNGKFS 718
DB 315 MOKYETTLAAGISQPRNRYNGNWTNSVSTNQTQNLERKRAFSGGVW-----YVDRAG 368
OY 719 QEALLA-----GVAV-----HKT-----VADNLVPMRGY-----RQRY 747

```

Db 369 IDARLGAFLAEGRKIPGSAYDLGNSHATMLTASMKRQLNNVLPENGHLDKIGTTL 428
Qy 748 SLEVGSSGLVSDANMAIARAGISGVSPFGDNAGSNRAHQMTGCI---QAGYIMSDNFN 803
Db 429 GTFLSSTALIRTS---ARAG-----YFPTPNKKLGTFFIIRGAGGYVARDNA 473
Qy 804 HVPYRLRFPAGGDSIRNGYAHDSLSPIISDKGYLTGGVLAAGTAEYNEFPKDLRLAVFG 863
Db 474 DVPSGLMFRSGGASSVGYELDSIGLAGPNGSVLPERALLVSGLEYQLPFTRTLSCGAVFH 533
Qy 864 DIGNAYDKGFTNDKIGAGVGRWASPVGVQVADVATGVKEGNPIKLHFIIGTPE 919
Db 534 DMGDAANFRMKLKHSGSLGVRFWFSPLAPPSFDIAYGHSDK--KIRWHISLGTFR 587
RESULT 25
AA74861
ID AAY74861 standard; Protein: 615 AA.
AC AAY74861;
XX 21-MAR-2000 (first entry)
DE Neisseria meningitidis ORF 286 protein sequence SEQ ID NO:1196.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX 11-NOV-1999.
PD 30-APR-1999; 99WO-0509346.
PE
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR MPI: 2000-062150/05.
DR N-PSDB; AA253623.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2: Page 667; 1453pp; English.
XX
CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC presence of Neisseria bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.
XX
SQ Sequence 615 AA;
Query Match 4.5%; Score 215; DB 21; Length 615;
Best Local Similarity 22.18; Pred. No. 2,4e-06;
Matches 132; Conservative 73; Mismatches 225; Indels 166; Gaps 20;
Qy 462 PVIEVELTDCILMDISPIEFSSASMLIODKLNLVAKARHLYDMPDDRVLA1NHDDGYNR 521
Db 48 PDTESVKLKPFPVILIDQSEIKDMVEHLPLITQDQEEVLDEKQTFLEAPADNKT 107
Qy 522 SI--LGRISDAVA-----VARAILPDESENEVIDLPERTALAN 558
Db 108 MLRSKGYFSSKVSLETKDGAVTHITPGPRKIANVGAIIIGDILSDCNLEYYRNALEN 167
Qy 559 RKTPA-----DYQSKR-----VPLVYVASKP 582
Db 168 WQDPVSGDQDSDSWENSKTSVIGAVTRKAYPLAKLGNTOAAVNPDTATADLVNVDVSGRP 227
Qy 583 RDGIGLGWGSDFGTR-----LYTKFEHNLINRDGYOAGEL- 619
Db 228 ----IARGDEFITTTQRPQIYVSGLARFQGMPIYDLDLDFQAL-EQNGHTSGASVQ 282
Qy 620 ----RLSEDKKGVRLYATKPLSHPLNDQLRAT---LG-----YQGEVFGHSTNG 661
Db 283 ADPFRLGQDRVPVAVSVTEVAKRHKLENGIRLDSYGLGKLAIDYNNLFNKGIGSVVWD 342
Qy 662 FDLSTRTLEHEISRSIIIONGW---NRTYSLRRLDKLKTQAPPETWODLPVDFVNGKPS 718
Db 343 MDKYETTLAAGISQPRNRYGNWYTSNVSYNSTQNLKRAFSGGVW-----YVRDRAG 396
Qy 719 QEALLA-----GVAY-----VADNLVNPKNGY-----RGRY 747
Db 397 IDARLGAFLAEGRKIPGSAYDLGNSHATMLTASMKRQLNNVLPENGHLDKIGTTL 456
Qy 748 SLEVGSSGLVSDANMAIARAGISGVSPFGDNAGSNRAHQMTGCI---QAGYIMSDNFN 803
Db 457 GTFLSSTALIRTS---ARAG-----YFPTPNKKLGTFFIIRGAGGYVARDNA 501
Qy 804 HVPYRLRFPAGGDSIRNGYAHDSLSPIISDKGYLTGGVLAAGTAEYNEFPKDLRLAVFG 863
Db 502 DVPSGLMFRSGGASSVGYELDSIGLAGPNGSVLPERALLVSGLEYQLPFTRTLSCGAVFH 561
Qy 864 DIGNAYDKGFTNDKIGAGVGRWASPVGVQVADVATGVKEGNPIKLHFIIGTPE 919
Db 562 DMGDAANFRMKLKHSGSLGVRFWFSPLAPPSFDIAYGHSDK--KIRWHISLGTFR 615
RESULT 26
AAB46311
ID AAB46311 standard; Protein: 916 AA.
XX
AC AAB46311;
XX
DT 05-APR-2001 (first entry)
XX
DE H. pylori HPS120 protein.
XX
KW Microbial infection; antibacterial; Helicobacter pylori infection;
KW vaccine; screening.
XX
PN WO200073502-A2.
XX
PD 07-DEC-2000.
XX
PF 31-MAY-2000; 2000WO-EP05024.
XX
PR 31-MAY-1999; 99DE-1024965.
PR 17-JUN-1999; 99DE-1027740.
PR 21-JUL-1999; 99DE-1034029.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (CREA-) CREATOGEN GMBH.
 XX
 PI Apfel H, Fuchs TM, Gibbs CP, Hueck CJ, Meyer TF;
 DR WPI; 2001-049948/06.
 DR N-PSDB; AAF25588.
 XX
 PT Preparing an agent for diagnosis or control of microbial infection,
 PT useful particularly against Helicobacter, based on identification of
 XX essential genes in defective mutants.
 XX
 PS Claim 37; Page 236-239; 366pp; German.
 XX
 CC This invention describes a novel preparation of an agent (A) for
 CC detection, prevention and/or treatment of microbial infection by:
 CC (i) identifying essential genes (I) and corresponding polypeptides
 CC (II); (ii) identifying compounds that are directed against (II) and
 CC inactivate the microbe; (iii) testing these for suitability for use; and
 CC (iv) formulating selected (A). Identifying essential genes (I) comprises
 CC preparation of gene-deficient microorganisms by conditional antisense
 CC inhibition (CAI) and/or subtractive recombination mutagenesis (SRM),
 CC then determining viability and/or survival of the deficient organisms.
 CC The products of the invention have antibacterial activity. (A) (which may
 CC be a nucleic acid (Ia), vector or host cell containing (Ia), derived
 CC polypeptide (IIa), or fragments, (IIa)-specific antibodies or their
 CC fragments or an inhibitor of (IIa)) are particularly used for diagnosis,
 CC treatment or prevention of infection by Helicobacter pylori. Particularly
 CC (Ia) and (IIa) are used in DNA, subunit or live vaccines. The method
 CC identifies essential genes, including those that have homologs in other
 CC species, so identified (A) should have a broad spectrum of activity. Many
 CC gene-deficient cells can be screened quickly, in an automated process,
 CC and the identified genes can be used for screening without purification.
 XX
 SO Sequence 916 AA:
 4.4%; Score 206.5; DB 22; Length 916;
 Query Match Best Local Similarity 19.3%; Pred. No. 1.6e-05;
 Matches 202; Conservative 143; Mismatches 318; Indels 383; Gaps 51;

QY 459 T-LEPVIETVELTGLDMD---ISPIEFSASMLT-----QDKLNLVAKARHLTYMP 506
 DB 379 NGLVAVIYRIEVDKMYINDVIISGNQRTSDRIIRRELLGPKKYNLTCLR----- 430
 QY 507 DDRVLAINHDDGVNRSILIGRISDAVSAVARAILPDESENEVDLPERIALNRKTPADY 566
 DB 431 -----NSENSLR--LGFEISKVIEEKRY-----NSSLMDL----- 459
 QY 567 QSKVPLVVEVADSKPRDGI--GLMGSDTGTRLVTFEHLNIRDGYOAGAEIRLSED 624
 DB 460 -----LVSEBGRGLOFGIGYSGGLM-----NCSYSERNLFGTQGSMLAN 506
 QY 625 --KKGVIYATKP-----LSHPLMDQLRATLGTQOEVPFGHSTNGFDLSTRLL- 669
 DB 507 IATGGGRSPGMPKAGRMFAGNLSTNP-----RLF-----DSWYSTIMLY 549
 QY 670 -EHEISRIIIONG-----WNRYY-SLRRLDKLTKQAPPETWQDLRVDFVNG-- 715
 DB 550 ADYRISTYQYIOGGGFGVNVGRMLGNRTHTVSLGYNLVTKLLGFSPLLYNRYSSVNEV 609
 QY 716 KPQGEALLAGVAHRTVA-----DNLVNP-----MRGYROR-----YSLEV 751
 DB 610 SPRGCTSPASVIYINRLSGKTPLPQESCSSPGAITTSPEIRGIMDROHTPTSSFTLDV 669
 QY 752 G-----SSGLVSDANNAIIRAGISGYVSFGD-----NAVSSNRAHOMTGG----- 791
 DB 670 SYDNTDTPYPRNGVIRS-----SYATMSGILPSSGTLNMGILGNVANTKYVKFPAVH 724
 QY 792 -----IAGYIMSDNF--NHVPYRLRFAGGQOSIRGAHDSLPSIDKGY 835
 DB 725 HLGKYLILDLIARKKTGGGIFRYNTDDYLPNSTFYMGVTVYGRNGSVTPBKDEGL 784
 QY 836 LTGGQVLAVGTAEYNEFMK--DLRLAVFGDIG-----N 867
 DB 785 WLGGDGIFFASTELSYGLVLAAKMRLAMFPDFGLTERTKTPRGSFFYNAPVTANPFDYG 844
 QY 868 AYDKGFTNDI-KIGAGYGVNRASPVG 892
 DB 845 VIGAGFERATWRASTGLOIEMISPMG 870

RESULT 27
 ABG25558
 ID ABG25558 standard; Protein; 371 AA.
 AC XX
 XX ABG25558;
 DT XX
 XX 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #25549.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PR 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PA (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS89745.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in


```
Db 315 VKEGIQYRISDL-----IEIDNPVPLK--TLEKALKVKRKDVFNIEHLRADQ 362
Oy 413 DL---IATRYENMVTLEVPPEEIOONDQVSEQSSSRTEPAQVDEST-LEPIETVE 468
Db 363 ILKTEIADKGYAFA---VVKPDLK-----DEKNGLVKVIYRIE 398
Oy 469 LTGILMDISPIEFASNLIDOKMLNVAAKARHLYDMPDR---VLAINHDDGVNRSILG 525
Db 399 VGD---MVHINDVIISGNQRTSDRI-----IRRELLGPKDKYNLTKLNSNSLR--LG 449
Oy 526 RISDAVSAVARAILPDESENEVIDLPERTALANKTPADVQSKVPLVYFVADSKPRDG 585
Db 450 FFSKVKIEEKRV-----NSLMDL-----LVSEEGRTG 478
Oy 586 QI--GLGWSDTGTRLVYKFEHNLINRDGYOAGAEIRLSED--KKGYKLYATKP----- 635
Db 479 QLOFGIGYGYGLML-----NGSVSEKRLCTGQSMILYANIATGGKSYFGMPKAGARM 534
Oy 636 -----LSHPLNDQLRATLGYOEVEFGHSTNGFDLSTRTL--EHEISRTIQNG----- 682
Db 535 FAGNLSTLNP-----RIF-----DSWYSTINLYADYRISYQIQGGGFGV 577
Oy 683 -----NNRTY--SLRYRLDKLKTQAPETMODLPVDVFN--GKPSQEALLAGVAVHK----- 730
Db 578 VGRMLGNRTHVSLGYNLNTVTKLLGFSSPLYNRYSSVNEVASPQOSTPASVITIRLSGG 637
Oy 731 ---TVADNLVNP-----MRGYRQR-----YSLEVG-----SSGLVSDA 760
Db 638 RPLVPESSCSFGAITSPEIKGIMWRDHTPTTSSEFTLDVSTDNDDYFPPRGVLES- 696
Oy 761 NMAIAGISGYVSPGD-----NAVGSNRHOMTGG-----IOAG 795
Db 697 -----SATNMSGLPSSSTGLSMWGLGNVNRNTKYGKFAAYHHLQKLLIDLIARFKTQGG 752
Oy 796 YWSDNF--NHVEYRLREFAGDOSIRGYAHDSLSPISDKGYLGGVLAAGTAETVYEFM 854
Db 753 YIFRYNTDDYLPINSTFYMGVITYVNGFNGSITTPKDEGLMLGGGIFTASTELSYGL 812
Oy 855 K--DLRLAVFGDIG-----NAYDKGFTNDT--KIGAGVCY 885
Db 813 KAAMKRLMFEDEGFLTEKTPRGSEFYNAPTTANFKYGVVAGAFERATMRASIGLOI 872
Oy 886 RMASPVG 892
Db 873 EMISPMG 879

RESULT 29
AAV17183
ID AAV17183 standard; Protein: 925 AA.
XX
AC AAV17183:
XX
DT 03-AUG-1999 (first entry)
XX
DE H. pylori outer membrane polypeptide.
XX
KW Outer membrane polypeptide; OMP; vaccine; H. pylori infection; humoral;
XX cellular immune response.
XX
OS Helicobacter pylori.
XX
PN MO9921959-AZ.
XX
PD 06-MAY-1999.
XX
PF 28-OCT-1998; 98WO-US22883.
XX
PR 17-DEC-1997; 97US-0993001.
XX 28-OCT-1997; 97US-0959131.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
```

```
XX Alm RA, Ellis RW, Guild BC, Noonan BM, Smith D;
PI WPI: 1999-326698/27.
DR N-PSDB: AAX75802.
XX
PT Cellular vaccine against Helicobacter pylori
PS Claim 7; Page 243-247; 352pp; English.
XX
CC The invention relates to a vaccine for preventing or treating infections
CC by Helicobacter pylori. The vaccine contains at least one isolated
CC H. pylori polypeptide, or its fragments, in a carrier, where the
CC carrier is a Salmomella, Vibrio cholerae or Shigella vector containing a
CC nucleic acid encoding the H. pylori polypeptide. The vaccines induce
CC humoral and cellular immune responses. The vaccines are used to treat or
CC prevent infections by H. pylori. Sequences AAX75779 to AAX75837 represent
CC nucleic acid sequences encoding H. pylori outer membrane polypeptides
CC (OMPs) AAV17160 to AAV17218.
XX
SQ Sequence 925 AA;
Query Match 4.1%; Score 194.5; DB 20; Length 925;
Best Local Similarity 19.9%; Pred. No. 0.00011;
Matches 192; Conservative 131; Mismatches 311; Indels 333; Gaps 49;
Oy 116 TPISLEELFQEST-----ENGINPNDIYIP-EYQ-----GEOPNSEYVY 153
Db 56 TPKEAQKNEAQNETSQNQPKREKVKYSISYVGLSYSDMLANELAKTRVGDWVDSKIID 115
Oy 154 PPTLEPKPKGLIKRLVYARLPNDGVNKKVPRLAKAYQSSQSG--ETSAGSSHQTEPYAN 211
Db 116 TAYALFENQGFYKDVYA--TEENG-----LEFHDEKARIAGVEIKGYEKEKGGLKSQ 169
Oy 212 IKALEDITOESAMDLDGSIPLRQAL--VAARAVGYDDLDLIRNSIGEVNVIINDLG 270
Db 170 MGAKGGTPEQKLE-----HAKTALKTALKEQGY-----GSV----- 203
Oy 271 EPIYIDYRAVEVRGEGADKAFTTVAD--EVLPLIGDVPHGKYEYKKNLIENASAEN- 326
Db 204 -----VEVTERKYESGALLIVFDVNRGDSIYIKQSIYESSDKRAKRVIESLSANKQ 254
Oy 327 -----GYFQGR-----WDRSVD--VILPD--NTADVSLIYD 354
Db 255 RDEMGMMGMLNDGKRLDLEYSRLIODVYMRGGLDAHISSPFKTDPSTHAKLHYK 314
Oy 355 T--GTQYRFDEVVFTIDPKNTLTDPPKLPYKRELLBOLLVNMGEAYNLQAVALSIN 412
Db 315 VKEGIQYRISDL-----IEIDNPVPLK--TLEKALKVKRKDVFNIEHLRADQ 362
Oy 413 DL---IATRYENMVTLEVPPEEIOONDQVSEQSSSRTEPAQVDEST-LEPIETVE 468
Db 363 ILKTEIADKGYAFA---VVKPDLK-----DEKNGLVKVIYRIE 398
Oy 469 LTGILMDISPIEFASNLIDOKMLNVAAKARHLYDMPDR---VLAINHDDGVNRSILG 525
Db 399 VGD---MVHINDVIISGNQRTSDRI-----IRRELLGPKDKYNLTKLNSNSLR--LG 449
Oy 526 RISDAVSAVARAILPDESENEVIDLPERTALANKTPADVQSKVPLVYFVADSKPRDG 585
Db 450 FFSKVKIEEKRV-----NSLMDL-----LVSEEGRTG 478
Oy 586 QI--GLGWSDTGTRLVYKFEHNLINRDGYOAGAEIRLSED--KKGYKLYATKP----- 635
Db 479 QLOFGIGYGYGLML-----NGSVSEKRLCTGQSMILYANIATGGKSYFGMPKAGARM 534
Oy 636 -----LSHPLNDQLRATLGYOEVEFGHSTNGFDLSTRTL--EHEISRTIQNG----- 682
Db 535 FAGNLSTLNP-----RIF-----DSWYSTINLYADYRISYQIQGGGFGV 577
Oy 683 -----NNRTY--SLRYRLDKLKTQAPETMODLPVDVFN--GKPSQEALLAGVAVHK----- 730
Db 578 VGRMLGNRTHVSLGYNLNTVTKLLGFSSPLYNRYSSVNEVASPQOSTPASVITIRLSGG 637
```


QY	329	FDRMLDRLSDVYLTPRNTADVSLYDTGTQVREDEVVFETIDPKTNOLLTDPDKLPVRE	368
Db	242	VREFLIDAKLMLNEDKNRIFLVEISLHEGEVRFQGTQFL-----GNLTQTQA	288
QY	389	LEFQLLTVMNGEYVNLQAVRALSNLNLTR-----YNNMNVN--TELVFPEREDIONDY	440
Db	289	ELFALLLKFAEGEFS-QAMLEQTTNNISTKFGDDGYAAQIRPVTRI-----NDE-	337
QY	441	SFEQSSSSFTPEPAVDDESTLVEFTELVTGILMDISPIEFASNLIDOKLNLVAKAR	500
Db	338	-----SKTVVEYXIDPVHPVY-----VRIINFQENFTQDE--VLREMR	376
QY	501	HLVYMDDDVNLINMDDDGNRSILGRISDAVSAVARAILPDSESENEVIDLPERTALNRK	560
Db	377	QL-----ECLASNOKIQLSR-----ARLMKRGFFKHVYVD--TRPVNSP	415
QY	561	TPADVQSKKVPVLYFVASDKPR-DGQIGLGMGSDTGRRLTKREHNLIRDGVQAGAL	619
Db	416	DOVDV-----FVVEEQRSGSSTILAAGYSQSGCVTTFQFDVQSQNNFMGTGHHVASF	466
QY	620	RLSEDKKGVKLYATRPILSHPLMDQLRALVLYGOAEVFGHSTNGEDLSTRLENEISRIID	679
Db	467	SRSETRREVYSLGTMTPY-----FTVNGVSQSLSGYRKTKYDNKNKINSVYLD	513
QY	680	NGGNMNTVSLARYRLD-----KL-----KTOAPPETMO	706
Db	514	STGGSLSLG--VPIDENORISFGLNADNTKLHGREFMGISNYQKOLMADGCKIYVDN--N	568
QY	707	DLVPDVFVNGKPSQEOALLGAVVHKTVDNLVNPARGYRQRTSLVEGSSGLVSDANNAIAR	766
Db	569	GIP-DFKHIDYTYNAILGM--NNSLDRPVFTPIQGNHSHVDLTG-----	610
QY	767	ADISGVYFSGDNAYGNSNAHQ-----MTGSIQAGYIWSDFNHNVPYRLRFEAG	814
Db	611	-----FGD-----KTHOKVYVYOGNIYRPFIKKSVLGRGAKLGGNNLPYENFVAG	656
QY	815	GDOSIRGAHDSLPISDKGYL-----GGQVLAAGTAE-----YNTEPMK	855
Db	657	GGVSGVRGIDQSSSLGPRS-QAVLTARRGQOTTLGEVGNALATFGSELLPLPFKGDWID	715
QY	856	DLRLAVFGDIGNAYD-KGFTNDT-----KITGAVY	883
Db	716	QVRPIYFIEGQVDFDTGMKQTDIDLPKQPDATAEQNKANKANRLLTQDQOLRTSAGV	775
QY	884	GVRMASPYGVQRVNDVATGV--KEGCAPIKLAHFIITGPF	919
Db	776	GATWYTPIGPLISYAKRPLNKKNDQDTQVOIGSYF	813
RESULT 34			
AAV44391			
ID	AAV44391	standard; Protein: 813 AA.	
XX	AAV44391:		
AC			
XX	DT	14-MAR-2000 (first entry)	
XX			
DE		M. catarrhalis (ATCC 43617) BASB027 polypeptide #2.	
XX			
KW		BASB027; OMP85; outer membrane protein; otitis media; treatment;	
XX		diagnosis; bacterial infection.	
OS		Moraxella catarrhalis.	
XX			
PN		W09963093-AA2.	
XX			
PD		09-DEC-1999.	
XX			
PF		31-MAY-1999; 99WO-EP03822.	
XX			
PR		03-JUN-1998; 98GB-0011945.	
XX			
RR		08-MAR-1999; 99GB-0005304.	

XX	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
PA	Vlnals-Bassols C;
PI	
XX	
DR	WP1; 2000-105700/09.
N-P	SDB; AA629551.
XX	
PT	Novel BASB027 polynucleotide and polypeptides from Moraxella
PT	catarrhalis useful for treating M. catarrhalis infection such as otitis
media	-
PS	
XX	
XX	Claim 1; Page 102-104; 109pp; English.
CC	The present sequence is BASB027 polypeptide, which shows significant
CC	homology to Neisseria meningitidis Omp85 outer membrane protein. It is
CC	encoded by DNA obtained from chromosomal DNA library of Moraxella
CC	catarrhalis strain MC2931 (ATCC 45617). BASB027 polynucleotide and
CC	polypeptide can be used for diagnosis and staging of disease, determining
CC	susceptibility to a disease and to prepare medicaments for treating M.
CC	catarrhalis infections, especially otitis media. The BASB027 DNA can be
CC	used as probe for screening of genetic mutations, serotype, taxonomic
CC	classification or identification. BASB027 agonists, antagonists and
CC	antibodies may be used to prevent and/or treat bacterial infections.
XX	
SQ	Sequence 813 AA:
Query Match	3.7%; Score 176.5; DB 21; Length 813;
Best Local Similarity	18.7%; Pred. No. 0.0016;
Matches 164;	Conservative 119; Mismatches 338; Indels 257; Gaps 36;
OY	165 IKRLVRLRNDGNNKPRLKAKFYOSSGSGETSAIGSSHQKTEPYANTRKALEDITQESA 224
Db	70 VVALVLTGNFSDVQYVHQSGRIITYOTTERPLLAETNFBGNRLIPBEGLQEGKNAGLAVG 129
OY	225 MDLNGSIPLRLQTALVAAR-AVGYYDIDISITRNSIG---EVDYIINDGEPPYIDYRA 279
Db	130 QPLKATVQMIEFTLNQYSOGYNTETTVAKQTMLDGNRVKLDMTFAE-GKPA---RV 184
OY	280 VEVRGGADDAKFTYADV-----PLLIGVFHHKKYETKKNLIENASAEH---GY 328
Db	185 VINIGNOHFSDADDLIDLAIKDKNINPLSKADRYOEKIYTS--LENLRAKYTLNMGF 241
OY	329 FCGRMILRSVDVITLRDNADVSLIIDTGQYRFEDVEVPFTDPKTNQLTTPDKLPVKRE 388
Db	242 VFETFKDAKLNIEMEDKNRIEVEISLHEEOYRFGTOFL-----GNLTYYQA 288
OY	389 LLEOLLTVMGAAYMLQAVRALSNDLIAIR-----YNNMVN--TEIVFPREOIQNOY 440
Db	289 ELEALLKFAREGFS-QAMLEOTTNNISTKEGDGYTAQIRPVTRI-----NDE- 337
OY	441 SFEQSSSRTEPAQVDESTLEPVIETVELTDGIIMDISPIEFSSNLIDOKLINLVAAKAR 500
Db	338 -----SRIVDEVXYIDVPHPVY-----VARINFGTGFKTODE-VLRRPMR 376
OY	501 HLHYMDPRDVIALINHDDGVNRSILGRISDAVSAAKAILPDSENEVYIDLERPALANRK 560
Db	377 QL-----GALASNOKIOLSR-----ARLMRTGFFGHVYVD--TRPVNSP 415
OY	561 TPADVYGSKKVPLEVFAVSDKPR-DGOIGIGMGSDPTGRILTYTKRENHLINDGYQAQAEI 619
Db	416 DOYVN-----FVVEEOPSGSSTIAAYSOSGGVTGFQVDSQNNEFMCTGHNVASF 466
OY	620 RLSEBDKGVKLKYATKPLSHPLINDOLRAVLGQAQEVFGHSTGFDLSRTLHEHSRSIIQ 679
Db	467 SRSEIREVLYSLGMNPY-----PTVNGVSOSLSGTRYRKTIYDKNKAINSYVLD 513
OY	680 NGGMNRTYSLARYLD-----KL-----KTQAPPETWQ 706
Db	514 SYGSGLSYG--YPIDENQRISFGINADNTKLHGGRFGMSNVKQLMADGKGILQYDN---N 568
OY	707 DLPDYFNVGKRSQALLAGVAHVHTVADNLVNPKRGTORYSTLEVSSGGLVSDANMAIAR 766

Db 569 GIP-DEFHDTTNAIGW---NTSLDRPVFPQGMHSVDTLVG----- 610
QY 767 AGISGVYSGFDNAGYNSRAHD-----MTGIGIQAGYIWSDNFNHVPYRLRFPAG 814
Db 611 -----FGD-----KTHQKVYQGNITRPFIKKSVLRGYALGIGNNLPFIENETIAG 656
QY 815 GDOSIRKYANDSLSPIDKGYLT-----GGQVLAVGTAE-----YNYEFMK 855
Db 657 GYGSVRCYDQSSILGPRS-QAYLTFARCGQQTLLGEVGNALATFGSELLPLPKGDMID 715
QY 856 DLRLAVGDIQGNAYD-KGFINDT-----KIGAGV 883
Db 716 QVRPVYIEBGGVFDTTGMDKQITDLTQFKDPQATAEQAKAANRPLLTDKOLRYSGAV 775
QY 884 GVRMASPVQGVNDVATGV---KEEGNPDKLHFFIGTPE 919
Db 776 GATWYTPIGPLSTISYAKPLKKKONDQTDYQFOIGSVF 813
RESULT 35
ABBS8843
ID ABBS8843 standard; Protein; 2748 AA.
XX
AC ABBS8843:
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 3321.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN MO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EM;
XX WPI: 2001-656860/75.
XX N-PSDB: ABL02946.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 3321; 21bp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC sequences (ABBS7737-ABBS72072).
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2748 AA;

Query Match 3.7%; Score 173.5; DB 22; Length 2748;
Best Local Similarity 19.6%; Pred. No. 0.014;
Matches 177; Conservative 131; Mismatches 312; Indels 285; Gaps 42;

QY 9 NRSEFPVALAAYLPLMTSOALAOONPANIIINHYPADHTAI-----NQAKGNPVLTPPE 64
Db 1217 SRAIEPLPLAP-----PAVEIEVVPGLGAEILLSEPEIEVVSFSAIALPE 1261
QY 65 QIQARLNAAGIN---AKPOSALDVNFDQSPISIRIGOSPPLGLDMVIEETPLSL 120
Db 1262 KPEPMENATEIEIEBMSVPTSEIIVPTTEDLHQAEVSEVETARP-----VSAVEEVKPI-- 1315
QY 121 EELFAQESTENGINPNDIYPEYQGE-QPNSEVVPPTLEPKPGILKLYARLNDCYK 179
Db 1316 -----EDSTAIDKSPDISPIEBEHSQSEKETPESEVE---VGIDSPNADVITTSVEV 1366
QY 180 VPRLKAIFYOSSOGSETSAIGSHQKTEP-----YANIKAALEDITQ----- 221
Db 1367 IP-----TSTAIVTTSQCASTEMPEPKYICETAPPAABEIEQNETVETLSYADTT 1418
QY 222 --ESAMDLNGSIPRLROTALVAARAVGYDIDLSIIRN-----SIGEVDTI----- 266
Db 1419 LPEEIVSQDGNLEASQTA-----SIDVTFQANLNMEKSGCEVAVDRESKED 1468
QY 267 HDLGEPIYIDYRA-VEV-----REBGADDKAFTTVADEVP 300
Db 1469 ETLMEPIAVDVAVEVVEVPPQAGVELGYSEPEEATPATNLTSGENPPEEV-SELVDVEP 1527
QY 301 LLIGDVFHGHKYEETKKNLIEASAEHGFQGRMLDRSY-----DVLIPDNADY 349
Db 1528 RNPHEEIHDTDSFEKDSOKSSDAVALHE--DWTKEEKVDPPQVAEVAAPDITITDPSLV 1565
QY 350 SLIYDT-----GTQYRPD-----EVVFTIDRKTNLTQTPD----- 381
Db 1586 ELSPDTQAMAGINLAPDILIEPESIVIEQVPVVKETPIVESDPEADIDHTALLEPSVS 1645
QY 382 -----KLVPKRELEQLITVNMGEAYVLQAVRALS-NDLIATRYFNKVN----- 424
Db 1646 AEYODKVESAOQVSEKNSVEEDSLKYSIGERGEIISMDPTPADAVSEQNFRVEDPTTS 1705
QY 425 -----TEIVPEREQIQNDQVSFEQSSSRTEPAQVDESTEPIYE--TVBLTGIGILMDI 477
Db 1706 TESTAVEVIASESINTIVEPVSAQSPF-----IEDSTPEEVTQCSAVNSTUTEVASA 1758
QY 478 SPIEFSASNLIODKLANLV-----AKARHLVDMDDRVLAINHDDGVNRSILGRISDAV 531
Db 1759 APIQ-----VDQKIQPVHVELVQSIEASDQESPKD--TAQVSDSDVDAIILQVSETD 1810
QY 532 SAVARAILP-----DESENE-----VIDLPERTALNKRTPADVYOS--KKVP----- 572
Db 1811 SIVVEPSSPVKQSEDSQEGPPGSAQNVLDV-----GPVEMDQSAPEVEPETDPVA 1860
QY 573 -----LVYFVASDKPRQOQIGLQSGSDGTGRLVYKFEHNLINRQCYQAGAEURL 621
Db 1861 VEPSVPVEEVEVIEIOTAPKPEISDSVERTLSLQNTDLVAEAS--VDTPAENMGPIGT 1918
QY 622 SEDKKGKLYATKPLSHPLNDQLRATLG-----YQGEVFGSTGDFLSTRTEHEHSRS 676
Db 1919 SELEKELLYVT-----TIDQPSAVLEPLVLICQAE-----STN-----LDHELPEIN 1961
QY 677 IIONGNRTYSLRYRDLKLKTQAPPE-TWODLPEVDFVNGKPSGEALLAGAVAKHTYADN 735
Db 1962 LVPGQ-----QSVPASLPTNLTITQTPTD-LETOPSMELLITQ-----TEYVDAE 2004
QY 736 LVNPM 740
Db 2005 MKRPV 2009

RESULT 36
AAR53755
ID AAR53755 standard; Protein; 797 AA.
XX
AC AAR53755;
XX
DT 08-DEC-1994 (first entry)

```

XX DE H. Influenzae b Eagen D15 sequence.
XX XX Vaccine: passive immunization; vector; antiserum; diagnosis; D15;
XX KW OMP; outer membrane protein; Hib.
XX XX
XX OS Haemophilus influenzae type b Eagen strain.
XX XX
XX PN W09412641-A.
XX XX
XX PD 09-JUN-1994.
XX XX
XX PF 23-NOV-1993; 93WO-CA00501.
XX XX
XX PR 23-NOV-1992; 92GB-0024584.
XX PA (CONN-) CONNAUGHT LAB LTD.
XX PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
XX PI Yang Y;
XX DR WPI: 1994-200269/24.
XX DR N-PSDB: AA066199.
XX XX
XX PT Nucleic acid encoding D15 outer membrane protein - esp. of
XX PT Haemophilus influenzae, and related proteins, vectors, antisera
XX PT etc. useful in vaccines, for diagnosis and for passive
XX PT immunisation.
XX PS
XX PS Disclosure: Fig. 1B; 161pp; English.
XX CC
XX CC Outer membrane protein (OMP) D15 genes were isolated by screening
XX CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
XX CC Eagan and Minn A, and the non-typeable (NTH) strains SB33 and PAK
XX CC 12085. Nucleotide sequences were determined for the D15 genes
XX CC (AA066198-202) and the corresponding aa sequences were derived
XX CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
XX CC free of other antigens and lipooligosaccharides, by recombinant DNA
XX CC methods using the isolated genes.
XX CC
XX SQ Sequence 797 AA;
XX
XX Query Match 3.6%; Score 171.5; DB 15; Length 797;
XX Best Local Similarity 19.8%; Pred. No. 0.0034;
XX Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;
XX
XX QY 263 DVIITDCEPYIYDRA-----VEVRGECA-DKCAFTTVADEVPLLIGDFHNGKETK 315
XX | | : : : | : : : | : : : | : : : |
XX | | : : : | : : : | : : : | : : : |
XX DB 72 DVKAHQEDVLAVSVYAKSIISDVKIKGNSVPTKALKNLDANGCFKVDVL----IREK 127
XX
XX QY 316 KNLIENASAEHGYFDGRMLDRSVDI---LPDNTADVSLIYDTGYQRFDEVVFTIDPK 372
XX | | : : : | : : : | : : : | : : : |
XX | | : : : | : : : | : : : | : : : |
XX DB 128 LNEFAKSYKEIHASVGRY-NATVEPIVNTLPNNRAEILIQINEDDKAKIASLTF-----K 181
XX
XX QY 373 TQNLTTDPDKLPVKRNLLEQLLTVNMGAY-----NLQAVRALSNDLIATRYFMVNT 425
XX | : : : | : : : | : : : | : : : |
XX | : : : | : : : | : : : | : : : |
XX DB 182 GNE-SVSSSTIQEQMELQPDSSWMKLMGNKPFEGAQEFKDLQSR-----DYLLNGYA 232
XX
XX QY 426 EIVPEREQIONDOVSFEQSSSRTEPAQVD-----ESTLEPIVETVELTDG 472
XX | : : : | : : : | : : : | : : : |
XX | : : : | : : : | : : : | : : : |
XX DB 233 KAOITTKTDVQINDEKTKVNTIDVNEGLOVDLRSARIICNLGMSAELEPLLSAHLNDT 292
XX
XX QY 473 I-LMDISPIEFSASNLIDDKNLINVAKAARHLYDMPD-----DRVLAINDDDGNRSIL--- 524
XX | : : : | : : : | : : : | : : : |
XX | : : : | : : : | : : : | : : : |
XX DB 293 FRSDIADYF-----NAKAKAGERGYSATVNSVPDPDANTLTALTIVDAGRRLTVRQ 348
XX
XX QY 525 ---GRISDAVASAVARAILPDES---ENEVIDLP---ERT---ALANRRTPAD----- 564
XX | : : : | : : : | : : : | : : : |
XX | : : : | : : : | : : : | : : : |
XX DB 349 LRFGNTVSAOSTLRQEMRQEGTMVNSQLVELGKIRLDRGFEFTVENRDIRDPIGNSDE 408
XX
XX QY 565 ---VYQSKKVPFLVYFVASDKPRDGOI--GLGWSGDTGRLVTKF-EHNLTI----- 608
XX | : : : | : : : | : : : | : : : |
XX | : : : | : : : | : : : | : : : |
XX DB 409 VDVVYKVK-----ERNTGSINFGIGYGTESGISYQASVYKODMFLGTGAIVSIAG 457

```

```

QY 609 NRDGYOAGAEILRISF---DKKGVKLYATKPLSHPLMDQLRALIGYQOEYFGHS-TNGFDL 664
XX | : : : | : : : | : : : | : : : |
XX DB 458 TKNDYGTSVNLGTYEPEYFKKDGVSLOGNVFFENYDNKSDTSSNRYKRTYGGSNVTGLGFFV 517
XX
XX QY 665 STRT-----LEH---EISRSIIQNGGMNPT---YSLRPLDKTKQAPPETWQDLPVDV 713
XX | : : : | : : : | : : : | : : : |
XX DB 518 NENNSTYVGLGHTYKISNPALE---YNRNLYIQSKKFGNGIKTN-----DFDPS 565
XX
XX QY 714 NGKPSQEALLAGVAHKTVAADNLVNPMPRGYRORYSLEVGSSGLV---SDANNAIARAI 769
XX | : : : | : : : | : : : | : : : |
XX DB 566 FGV-----NYSNLNRCGYFPPTKGVKASLGGVTTIPGSDNKYKLSADY 607
XX
XX QY 770 SGVYSRGDNAYSNRH--QMTGGIOAGYIWSNFMHVVYRLRFFFGQDSIRGAYHDSL 827
XX | : : : | : : : | : : : | : : : |
XX DB 608 QGFYPL-----DRDHLWVSAKASAGYANGFGNKRPLPEYQYTAGGIGSLRGFAVYSI 660
XX
XX QY 828 SP-----ISDKGYLTGGOVLAVGTAE--VYEFMKD-----LRLAVF 862
XX | : : : | : : : | : : : | : : : |
XX DB 661 GPRAIYAEYNGSGTGFKKISD--VIGNMAIATASAEILYPTFPVSDKQNTVTSIF 718
XX
XX QY 863 GDIGNAY-----DKGFTNDRKIGAGVGYRMASPVGOVRVDVATGV 902
XX | : : : | : : : | : : : | : : : |
XX DB 719 VDAASVWNTKMKKSDKNGLSDVLRPLPDYKSSRIRASTGVGFQWQSPIGPLVFSYAKPI 778
XX
XX QY 903 K--EEGNPTKLHFITGTPF 919
XX | : : : | : : : | : : : | : : : |
XX DB 779 KRYENDVBEQFOFISIGSF 797
XX
XX RESULT 37
XX AAR53756
XX ID AAR53756 standard; Protein; 797 AA.
XX
XX AC AAR53756;
XX
XX DT 08-DEC-1994 (first entry)
XX
XX XX H. Influenzae b Minn A D15 sequence.
XX
XX DE Vaccine: passive immunization; vector; antiserum; diagnosis; D15;
XX KW OMP; outer membrane protein; NTHI.
XX KW
XX OS Haemophilus influenzae type b Minn A strain.
XX
XX PN W09412641-A.
XX PD 09-JUN-1994.
XX PD
XX PF 23-NOV-1993; 93WO-CA00501.
XX PF
XX PR 23-NOV-1992; 92GB-0024584.
XX PR
XX PA (CONN-) CONNAUGHT LAB LTD.
XX PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
XX PI Yang Y;
XX PS
XX PS Disclosure: Fig. 1C; 161pp; English.
XX CC
XX CC Outer membrane protein (OMP) D15 genes were isolated by screening
XX CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
XX CC Eagan and Minn A, and the non-typeable (NTH) strains SB33 and PAK
XX CC 12085. Nucleotide sequences were determined for the D15 genes
XX CC (AA066198-202) and the corresponding aa sequences were derived

```



```
DB 409 VDVVYKVK-----ERNNGSINFGIGYTESGISYQASVKQDNFLGTAANVSIAG 457
QY 609 NNDGQQAQAEFLSE---DKGKYLATKPLSHPLNDOLRALGTYQOEYFGHS-TNGFDL 664
DB 458 TKNDYGTSVNLGTYTEPYTFKDGVSLOGNVFPENYDMSKSDTSSNYKRTTYGSAVTLGFPV 517
QY 665 STRT-----LEH---EISRSIIIONGWNRT---YSLRYLRLDKTKQAPEPMODLPLDVY 713
DB 518 MNNSIYYVGLGHTYTKISNFALF---YNNRLYIQSMKFKNGIKTN-----DFDS 565
QY 714 NGKPSQEALLAGVAHVKTADNLVNPMPRGYRQRYSLVGVSSGLV---SDANMAIARAGI 769
DB 566 FGM-----NYSNLMRGYFPTKGVKASLGGVHTIPGSDNKKYKRLSADY 607
QY 770 SCVYSEFGDNAYGNSRAH--QMTGGIOAGYIWSDNFNHVPYRLRFPGAGQOSIRGVAHDSL 827
DB 608 QOEYPL-----DRDHLWVYSAKASAGANGFGNKRPLFPYQTYTAGGIGSLGFPYGS 660
QY 828 SP-----ISDKYLTGGQVLAAGTAE--YNYEPMKD-----LRAYF 862
DB 661 GPNATYAEYNGSGTGTFFKKISSD--VIGGNALATASPELIYPTFVSDKSQNTVATSLF 718
QY 863 GDIGNAY-----DKGFTNDTKIGAGVGVNMAAPVGVAVDVATGV 902
DB 719 VDAASVWMTKWKSDKNKGLSDVLRPLDYKSSRIASTGVGFQMSPIGLIVFSYAKPI 778
QY 903 K--EEGNPIKLAHFICTPP 919
DB 779 KYENDVDEOFQFSTIGSP 797
```

RESULT 39
AAR45002
ID AAR45002 standard: Protein: 1319 AA.

AC AAR45002;

DT 13-JUN-1994 (first entry)

XX Cellulose synthase operon, gene C product.

KW Bacterial: cellulose synthase: operon; gene A: gene B: gene C;
KW gene D: transcription vector; recombinant microorganism;
KW cellulose synthesis.

XX Acetobacter xylinum.

XX US5268274-A.

XX 07-DEC-1993.

XX 12-APR-1989; 89US-0337194.

XX 10-APR-1990; 90CA-2014264.

XX 11-APR-1990; 90IE-0001317.

XX 09-APR-1990; 90IL-0094053.

XX 12-APR-1990; 90NZ-0233312.

XX 04-APR-1990; 90WO-US01811.

XX (CETU) CETUS CORP.

XX Ben-Bassat A, Ben-Ziman M, Calhoun RD, Fear AL, Gelfand DH;

XX Meade JH, Tai R, Wong H;

XX WPI: 1993-404004/50.

XX N-PSDB: AAO53522.

XX Contiguous nucleic acid sequences - encoding bacterial cellulose
XX synthase
XX Claim 52; Fig 1 and Columns 53-72: 79pp; English.
XX The sequences given in AAR45000-03 represent the proteins encoded by
CC

CC the bacterial cellulose synthase operon. The sequence given in
CC AAR45004 is the begining of an open reading frame overlapping the end
CC of this operon. The bacterial cellulose synthase operon contains
CC four genes, genes A-D. The operon sequence may be used in a
CC transcription vector for the expression of the cellulose synthase
CC operon to increase cellulose production in a recombinant microorganism.
CC This system may be used as an important tool for exploring mechanisms
CC of cellulose synthesis and for enhancing production of cellulose.

XX Sequence 1319 AA:

Query Match 3.4%; Score 159.5; DB 14; Length 1319;

Best Local Similarity 19.2%; Pred. No. 0.047;

Matches 201; Conservative 131; Mismatches 360; Indels 357; Gaps 51;

```
QY 43 PAHDIAI-----NQAKAGNPVLLT-PEQIQARL--NAGINAKPQSQAL- 84
DB 164 PPHSLAVEYQTMAGVPAQDQARAGIAGVVAANPDQYRQALFPAQALTYNTSTRMGLT 223
QY 85 ---DVNFPDQSPF-SRIGOSPPLGID-MSVIEETPLSLEELFAQESTEMGINPDYI 139
DB 224 RLKDLOSFSQAEVEAAQAQSTROTLSWLPNPEPTQPLEQWLSA-----HPMD- 273
QY 140 PEYQGEQPNSEVVVPTLEPEKPLIKRLYARLFNDGVNKPRLKAKFYQSOSGETSAI 199
DB 274 ---TALREHMLHPGCGPPDKAGIARQ-----AGYQQLNAGRILAA 310
QY 200 GSSHQKTEPYANIKAALEDITQESAMDNGSIFRLROTALVAARANGYDIDSLINSTI 239
DB 311 EDSFQ-----SALOINS-----HDAD-----SL 328
QY 260 GEVDVITIHDLGEPVYIDYRAVEYRGEGADKAFITVADEVPLLIGVFHHKYEYKKNLI 319
DB 329 GGMGLVSMKQGDYA--EARKYFEEMAADPK-----TADRRKRPALAGAVSGEYASVROLI 362
QY 320 ENASAEHGYFFDGRW-----LDRSYDVI----- 341
DB 383 ---AAHQYTEAKQQLATLARQPGQYTGATLMLADLQIRSTGQIAAAEQEYRGLISREPN 438
QY 342 -----LPDNTADV-SLIYDTGYOY--RPEDEVYFTIDPKTNQTLTTPDKLPVK 366
DB 439 QIALMGLIARYDMAQNTAEKROLLSRVGPQYASQVEIEVSGLMAAASQTSARAKVSIL 498
QY 387 RELLEQL-----LTVNMGEVYNLQ-----AVRAL-----SND 413
DB 499 REAMAQAPRDPWVRINLANALQOQGDVAEGRWQPILANPYTAQDRQAGILTYGSGND 558
QY 414 LIAIRYFNWVNTETIVEPEREJOIONDOVSFEQSSSKTEPAQVDESTLEPIYETVELTDGI 473
DB 559 AMTRQLAGLSPADYSPAIRSIA-EEMETIKQDLASRLSWSNVPLIREALTQPPDTGAR 617
QY 474 LMDISPIEFSASNLIODKLNIVAKARHLVDMDDV-----LAINHDCGVRSITGRI 527
DB 618 GVAVADLFRQGDVHARMAALRISTRTIDLSPORLSTATEYMKISNPVAAAR-LIAPL 676
QY 528 SPAV-SAVARAILPDESE-----NEVIDLPERTALNKRKTPADVYSKPYLVFVASDK 581
DB 677 GCGTGSATGSALLPEQVQLQQLRMGISVAQSDLLNORGQQAQAYHILAPALADBEATS 736
QY 582 PR-----DQIGLIGMSDGTGRLVTKFEHNLINRDOYQAGELRUSEDKGV-KL 630
DB 737 PRLALRLYNGHKKPKKALEID-----LAVLRHNPQDLARQAAYQAAVSDHNSLATRL 791
QY 631 YATKPLSHPLNQLKATLGLYQOEVEFGHSTNGPFLSTRTELEHSRLIONGGMNRTYSLR 690
DB 792 AMDGVOESPMADARAWLAMVADQDGH-----GORTLE-----DIRRAYDLR 833
QY 691 YRU-----DKLTKQAPPET-----WQDLPVDFVNGKPSQEA----- 721
DB 834 LQOVBECTRAASGAGAQEDAL---APPSTNPPFRPRRGYGHOTELGAVTGSYAEASAAPD 890
QY 722 ---LLAGVAVH-KTVVADNL---VNPMPRGYRQRYSLVGVSSGLVSDANMAIA-----RAG 768
DB 722 ---LLAGVAVH-KTVVADNL---VNPMPRGYRQRYSLVGVSSGLVSDANMAIA-----RAG 768
```

DB 891 TSDOMLSSIAQITLTLEENLAPSTIDGLGFRSR-SGEHG-MGRLEANTEPIVGRPLQAG 948
QY 769 ISGVYSEFGDNAGSNRAHOMTGTIOAGYIMSDNFN-----HVPYRLRFAGDQDSIRGY- 822
DB 949 ASA-----LFSITPTMIMSGNLTGSDYDVF---RY--CTMGVQAYN 987
QY 823 AHSLSLPSIDKGYITGG---QVLAAGTAETVYEFMKDLRLAVFGDIGNAYDKFTNDTKI 879
DB 988 QYDS-----YTNACRDOORLACTAE-----AGFAPDVQF 1017
QY 880 G-----AGVGRNASPVGYRVDAVTGK 903
DB 1018 GNSWVRADVC---ASPIGPFITNVLGVE 1043
RESULT 40
ABG21666
ID ABG21666 standard; Protein: 1249 AA.
XX
AC ABG21666;
XX
XX 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #21657.
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX N-PSDB; AAS85853.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 20: SEQ ID No 52025; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1249 AA;
Query Match 3.38; Score 157.5; DB 22; Length 1249;
Best Local Similarity 18.28; Pred. No. 0.06;
Matches 114; Conservative 88; Mismatches 181; Indels 245; Gaps 27;
QY 466 TVELTGIMDISPIESASNLIDKLNLA----- 496
DB 54 TITVNGVLDVNVDDFEAFRTDKLELTSGNIADHNGVNSGVFDIHSSDYLMADLV 113
QY 497 -----AKARHLYD---MPDDRVLAINH---DDGVN-----RSILG-----RISDA 530
DB 114 NDRTWDTSKSNYGYGIYAMNSDGHLLTTNGNDVDNGTWMIDLCNNVVAATGNVYRIDA 173
QY 531 VSAVARAILPDESENEVIDPE--RTALANKRPADV---YQSK----- 570
DB 174 TGAGANA---DYKDKETIYVDVNSNATFSANKRADLGAYTYQAEQKGNFTACLPQMELEDD 230
QY 571 ---VPLVVFVASDKPRDQOIGLW--GSDT--GTRLVTKFEHNLINRDCYQAGAELRUSED 624
DB 231 YANPPLSIPSANTNI-----WNLGQDVYGTBL-TNSRHGLADNG----- 269
QY 625 KKGVLVATKPLSHPLNDQLRATLGYOEYVGHSTNGFDSLRTLEHRSIT-----QN 680
DB 270 -----AWYSY-----FGGNFG--DNCTINYDDVNGIMVGVDTKI 303
QY 681 GGMNRTYSLRYRLDKLTKTOAPPETMODLPVD-----FVNGKPSQEL---LAGVAVHKY 731
DB 304 DGNNAKWIVGAAGSLKAVDDQGTAYIYSSAHFANNVYFVDSLSLSYHFNNDLSATMNGT 363
QY 732 VADLVNPMR---GYRQRTSLEWSSGLVSDANMAIRAGISGYISFGDNAGSN----- 783
DB 364 YVDSNSTDAMWFGKAGYFKLGADAGYVT-----PYGSVSLFQSGDDYQLSNDKQVD 417
QY 784 -----RAHOMT-----GGIAGCYIMSDNFN 804
DB 418 GQSYDSMKREGSFVPEASACKMKVTLDTATYPIIDDDHKWVYLVGRWYGDGLGKE 477
QY 805 VPYRLRFAGGDSIRGYANDSLSPIS-----DKGY-----L 836
DB 478 MPFEYENFYAGGSSYVRGFSQNTIGPKAVYFPHQASNTDPDVEDYCATQDCAKDKCKSDA 537
QY 837 TGGQVLAAGTAETV-----NYEFMKDLRLAVFGDIGNAYDKGF-----TND 876
DB 538 VGGNALAVASLEFETTPPFLISDKYANSVRTSFEMDMCTVMDTNWSSQYSGYPDYSDPSN 597
QY 877 TKIGAGVGRWASPVGYGVYRVDAVTGVE 904
DB 598 IRMSAGIALQWMSPLGLPVFSYAOPFRK 625

Search completed: April 28, 2003, 16:21:39
Job time : 101 secs


```

Db      72 DVKAHQEGDVLVSVVAKSIISDVKIKGNSVITPEALQKQNDANGKVGDL-----IREK 127
Qy      316 KNIENASAEHGFDGRMLDRSVDI---LPDNTADVSLITDGTQVREDEVFETIDPK 372
      128 LNEFAKSVKEHYASVGR-NATVEPIVNTLPNNRAELIQINEDDKAKLASLTF-----K 181
Qy      373 TNOITTDPKLPVKRELEQLLVNMGAY-----NLQAVRALSNDLITRFENMVT 425
      182 GNE-SVSSSTLOEQMELQPSMWKLMGNKFEQAQFEKDLQAIR-----DYLLNGYA 232
Qy      426 EIVPEREQIQNDQVSESSSSRTPEAOVD-----ESTLEPIEYVELTDG 472
      233 KAQITKTDVOLNDEKTKVNTITDVNEGLQYDLRSARIIGNLGSASLEPLLSALHNDT 292
Qy      473 I-LMDISPIEFSASNLQDKLNLVAKARHLTMDP---DRVLAINHDDGVNRSIL--- 524
      293 FRRSDIADVE---NAIKALIGERGYNCTTVNSVDPDANKTLAITFVDAGRHLTVRQ 348
Qy      525 ---GRISDAVSAVARAILPDES---ENEVIDLP---ERT---ALANKRTPAD----- 564
      349 LRFEGNTVSADSTLRQEMNQEGTWNLSQVELGKIRLDRTEFEFVENRIDPINGSND 408
Qy      565 ---VYOSKVPVLYVVASDKPRDQI--GLMGSDTGTRLVTKF-EHNL----- 608
      409 VDVVYKVK-----ERNTGSINFGIGYTESGISYQTSIKQDNFLGTGAANSIAG 457
Qy      609 NRDGVOAGAEILRSE---DKKVKLYATKPLSHPLNDQLRATLGYOOEVFGHS-TNGFDL 664
      458 TKNDYGTSVNLGYTEPEFTKDGVSIGGNIFPENYDNKSDTSSNKRRTYGSNVTLGFPV 517
Qy      665 STRT-----LEH---EISRSIIQNGGNMRT---YSLRYRLDKLTKQAPPTMODLVDEV 713
      518 NENNSYVGLGHTYKNKISNFALE---YRNRLYIOSMKFKGNGIKTN-----DFDS 565
Qy      714 NGRPSQBALLAGVAHVKTADVNLVNPMRGVRQRYSLVSGSSGLV---SDANMAIARAGI 769
      566 FGM-----NYNSLNRGYFPTKGVKASLGSRVITIPGSDNKYKLSADV 607
Qy      770 SGVYSFGDNAYGNSRAHO--MTGTOAGYIMSDNHNHVPYRLRFAGDQSIIRGYAHDLS 827
      608 QGFYPL-----DRDHMVVYSAKASAGYANGFGNKRLPFYQTYAGIGSLRGFAGYSI 660
Qy      828 SP---ISDKGYLT-----GGQVLAVGTAE--YNYEFKMD-----LRLAVFEDIGNA 868
      661 GPNALYAEHNGCTENKISSDVIIGMAITTSABELIVPTPEVSDKSQNTVTRSLFPDASV 720
Qy      869 Y-----DKGFTNDTKIGAGVGVMASPVCGOVRVDVATGVK--DEG 906
      721 WNTKWSKDKNGLESKVLKLDLPDYGKSSRIRASTGVGFQMGSPIGPLVFSYAKPIKYEED 780
Qy      907 NPIKLHFEICTPF 919
      781 DVEQFPQFSGCF 793

```

RESULT 2

US-09-135-166-10

Sequence 10. Application US/09135166

Patent No. 6083743

GENERAL INFORMATION:

APPLICANT: CHONG, Pele

APPLICANT: THOMAS, Wayne

APPLICANT: YANG, Yan Ping

APPLICANT: LOOSMORE, Sheena

APPLICANT: SIA, Dwo Yuan Charles

APPLICANT: KLEIN, Michel

TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN

NUMBER OF SEQUENCES: 55

CORRESPONDENCE ADDRESS:

ADDRESSER: Slim & McBurney

STREET: 6TH Floor, 330 University Avenue

CITY: Toronto

STATE: Ontario

```

? COUNTRY: Canada
? ZIP: M5G 1R7
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent in Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/135,166
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/433,522
? FILING DATE: 12-SEP-1995
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: STEWART, Michael I
? REGISTRATION NUMBER: 24,973
? REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (416) 595-1155
? TELEFAX: (416) 595-1163
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 793 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US-09-135-166-10

```

```

Query Match      3.8%; Score 180.5; DB 3; Length 793;
Best local Similarity 20.1%; Pred. No. 1,7e-06;
Matches 159; Conservative 121; Mismatches 306; Indels 207; Gaps 39;

```

```

Qy      263 DVIIDHLEBPYIDYRA-----VEVRGEA-DDKAFYVADEVPLIGDVFHNGKETK 315
      72 DVKAHQEGDVLVSVVAKSIISDVKIKGNSVITPEALQKQNDANGKVGDL-----IREK 127
Qy      316 KNIENASAEHGFDGRMLDRSVDI---LPDNTADVSLITDGTQVREDEVFETIDPK 372
      128 LNEFAKSVKEHYASVGR-NATVEPIVNTLPNNRAELIQINEDDKAKLASLTF-----K 181
Qy      373 TNOITTDPKLPVKRELEQLLVNMGAY-----NLQAVRALSNDLITRFENMVT 425
      182 GNE-SVSSSTLOEQMELQPSMWKLMGNKFEQAQFEKDLQAIR-----DYLLNGYA 232
Qy      426 EIVPEREQIQNDQVSESSSSRTPEAOVD-----ESTLEPIEYVELTDG 472
      233 KAQITKTDVOLNDEKTKVNTITDVNEGLQYDLRSARIIGNLGSASLEPLLSALHNDT 292
Qy      473 I-LMDISPIEFSASNLQDKLNLVAKARHLTMDP---DRVLAINHDDGVNRSIL--- 524
      293 FRRSDIADVE---NAIKALIGERGYNCTTVNSVDPDANKTLAITFVDAGRHLTVRQ 348
Qy      525 ---GRISDAVSAVARAILPDES---ENEVIDLP---ERT---ALANKRTPAD----- 564
      349 LRFEGNTVSADSTLRQEMNQEGTWNLSQVELGKIRLDRTEFEFVENRIDPINGSND 408
Qy      565 ---VYOSKVPVLYVVASDKPRDQI--GLMGSDTGTRLVTKF-EHNL----- 608
      409 VDVVYKVK-----ERNTGSINFGIGYTESGISYQTSIKQDNFLGTGAANSIAG 457
Qy      609 NRDGVOAGAEILRSE---DKKVKLYATKPLSHPLNDQLRATLGYOOEVFGHS-TNGFDL 664
      458 TKNDYGTSVNLGYTEPEFTKDGVSIGGNIFPENYDNKSDTSSNKRRTYGSNVTLGFPV 517
Qy      665 STRT-----LEH---EISRSIIQNGGNMRT---YSLRYRLDKLTKQAPPTMODLVDEV 713
      518 NENNSYVGLGHTYKNKISNFALE---YRNRLYIOSMKFKGNGIKTN-----DFDS 565
Qy      714 NGRPSQBALLAGVAHVKTADVNLVNPMRGVRQRYSLVSGSSGLV---SDANMAIARAGI 769
      566 FGM-----NTNSLNRGYFPTKGVKASLGSRVITIPGSDNKYKLSADV 607

```


ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,522A
FILING DATE: 12-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-434 MIS:jfb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 792 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-433-522A-8

Query Match 3.8%; Score 178; DB 3; Length 792;

Best Local Similarity 20.2%; Pred. No. 2.8e-06;
Matches 160; Conservative 118; Mismatches 306; Indels 210; Gaps 39;

QY 263 DVIHIDGEPYIDYRA-----VEVRGEG-ADDAFTTVADEVPLLIGDVFHGKYEK 315
DB 72 DVKAHOGEDVLVSVAKSIISDVKIKNSIIPPEALKQNDANGFKYGDLL-----IREK 127
QY 316 KNIENASAEHGIDGKRLDRSVYI---LPDNTADVSLIYDTGTQYRFEDEVFEFTIDPK 372
DB 128 LNEFAQSVKEHYASVGRY-NATVEPIVNTLPNNRAEILIQINEDDKAKLASLTF-----K 181
QY 373 TNOITTPDKLPVARELLEQLITVNMGEAY-----NLQAVRALNSDLITRYFNMYNT 425
DB 182 GNE-SVSSSTLQDEMELQPDSSMKLMKKEGCAOFEKDLQAIR-----DYLNNGYA 232
QY 426 EIVFPEREQIONDOVSFEQSSSSSTPEPAOVD-----ESTLEPYIETVELTDG 472
DB 233 KAQITKADVOLNDEKTKVNTIIVNEGIGYDLRSARITLNGMSAELEPLSLALHNDT 292
QY 473 I-LMDISPIEFSSNLQDKLNLYAAKARHLYDMPD---DRVLAINHDDGVNMSIL--- 524
DB 293 FRNSDIADVE---NAIKAKLGERGYNTTVNSVPDDDKAKTLAIFVVDAGRRLTVHQ 348
QY 525 -----GRISDAVSAAVARAILPDES---ENEVIDLP-----ERT-----ALANKRTPAD----- 564
DB 349 LRFEGNVSADSTLRQMRQOEGTWYNSQLVELDKIRLDRKGFEFEYENRIDLPIINGSND 408
QY 565 ----VYQSKKVPPLVYFVASDKPRDGI---GLGWGSDGTGTRLVTKF-EHNLI----- 608
DB 409 VDVYVYKVK-----ERNTGSINFGIGYGTESGISVQASVKQDNPLCTGAASVAG 457
QY 609 NRDOYQAGAEILRSE---DKRGVLYATKPLSHPLNDLQRLATLQYQGFVGH-S-TNGDL 664
DB 458 TKNDYGTSVNLGTEPFTFDGYSVLGCVNFFENDSKSDTSSYKRTYTGNSVTLGFPV 517
QY 665 STRT-----LEH---EISRSIIIONGNNRT---YSLRYRLDKLTKQAAPETMODLPVDFV 713
DB 518 NENNSYVVGCHTYNKSINALE---YNRMILYIOSMKFKKGIGITN-----DDFS 565
QY 714 NGRPSEBALLAGVAHVKTADVNLVNPGRYRORYSLEVGSSGLV---SDANMAIARAGI 769
DB 566 FGV-----NYNSILNRGYFPTKGVKASLGGRTVITPGSNKKYKLSADV 607
QY 770 SCYYSFGDNAGYNSRAH---QMTGGIAGYIMSDNFNHPVRLRLRFAGGDSIRYAHDSL 827
DB 608 QGYPL-----DRDILMVYSAKASAGYANGFNKRRLPYQYTYTAGTIGSLRGFAVAGSI 660

QY 828 SP-----ISDKGYLTGGVLAAGTAE--YNTFEMKD-----LRLAVEGDIGN 867
DB 661 GPNALYOGONNKNKISSD--VIGGNAIATASAEILVPTPEFVSDKSQNTVRSLSFDAAS 718
QY 868 A-----DKGFTNDTKIGAGYGVWASPYGVQVRDVAATGVK--EE 905
DB 719 VVNTWKMSDKNGLESNYLKDLPDYGRKSSRTASTGVGFQWSPSGFVPSYAKPIKKYEN 778
QY 906 GNPRLHFEGITP 919
DB 779 DVEQFOFSGISGSF 792

RESULT 5

US-09-135-166-8

Sequence 8, Application US/09135166

Patent No. 6083743

GENERAL INFORMATION:

APPLICANT: CHONG, Pele

APPLICANT: THOMAS, Wayne

APPLICANT: YANG, Yan Ping

APPLICANT: LOOSMORE, Sheena

APPLICANT: SIA, Dwo Yuan Charles

APPLICANT: KLEIN, Michel

TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESS:

STREET: 6TH Floor, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/135,166

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/433,522

FILING DATE: 12-SEP-1995

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: STEWART, Michael I

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-829 MIS:jfb

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 792 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-135-166-8

Query Match 3.8%; Score 178; DB 3; Length 792;

Best Local Similarity 20.2%; Pred. No. 2.8e-06;
Matches 160; Conservative 118; Mismatches 306; Indels 210; Gaps 39;

QY 263 DVIHIDGEPYIDYRA-----VEVRGEG-ADDAFTTVADEVPLLIGDVFHGKYEK 315
DB 72 DVKAHOGEDVLVSVAKSIISDVKIKNSIIPPEALKQNDANGFKYGDLL-----IREK 127
QY 316 KNIENASAEHGIDGKRLDRSVYI---LPDNTADVSLIYDTGTQYRFEDEVFEFTIDPK 372
DB 128 LNEFAQSVKEHYASVGRY-NATVEPIVNTLPNNRAEILIQINEDDKAKLASLTF-----K 181

Db 661 GPMALYGGQNNKFNKISSD--VIGGNALATASABELIYPTPEVSDKSONTQVTSLEFVAAS 718
QY 868 AY-----DKGFTNDPKIGAGVGNRNASPVGOVVDVATVVK--EE 905
Db 719 VMMTKMKSDKNGLESNVLKDLPDYGKSSRTIRASTGVGFQWQSPGPPVFFSAKFIKRYEN 778
QY 906 GNPFKLHFETGTP 919
Db 779 DVEQFQFQSIGSE 792

RESULT 7
US-08-433-522A-2
Sequence 2, Application US/08433522A
Patent No. 6013514
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: THOMAS, Wayne
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: SIA, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & McBurney
STREET: 6TH Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,522A
FILING DATE: 12-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-433-522A-2

Query Match 3.6%; Score 171.5; DB 3; Length 797;
Best Local Similarity 19.8%; Pred. No. 1e-05;
Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

QY 263 DVIHDLGEPVYIDYRA-----VEVREGGA-DDKAFYVADEVPLLIGVFHNGKYEK 315
Db 72 DVKAHQEGDLVAVYAKSIISDVKIKGNSVITPEALKQVLNANGFVGVVL----IREK 127
QY 316 KNIENASAEHGTFDGRWLDKRSVDV---LPDNTADVSLIYGTQYRDEVDVFFITDPK 372
Db 128 LNEFAKSVKEHVASVGRY-NATVEPIVNTLPNNRAEITLIDINEDKAKLASLTF-----K 181
QY 373 TNOTLTDPPKLPVKRELLLEQLTVNNGEAY-----NLQAVRALSMDLATRYFNKVVNT 425
Db 182 GNE-SVSSSTLQOMELQDPDSMMKLMKNKEGQAFKEDLOSIR-----DYLLNNGYA 232
QY 426 EIVFPEREIONDOVSFEQSSSSSRTEPAQVD-----ESTLEPIETVELTLDG 472

Db 233 KAQITKTDVQINDEKTKVNVITIDVNEGLOYDLRSARITIGNLGMSAELEPLISALHMDT 292
QY 473 I-LMDISPIERSASNLIDDKLNLVAAKARHLXYDMPD-----DRVLINDDGVNRSIL-- 524
Db 293 FRRSADIAYE-----NATAKALGERGYGSATVNSVDFPDANKTLAILTVDAGRRLVRO 348
QY 525 ---GRISDAVSAVARAILPDES---ENEVIDLP---ERT---ALANRKT PAD----- 564
Db 349 LRFEGNTVSAADSTLRQERROEGTWNSQLVELKIRLDRGTGFEFTVENRIDPINGSND 408
QY 565 ---YYQSKKVPPLYVFNASDKPRDGI--GLMGSDPTGRILYKF--EHLI----- 608
Db 409 VDVIYKVK-----ERNYGSINFGIGYGESIGISYASVKQDNFLTGAAVSTAG 457
QY 609 NRDGQAGAEELRLE---DKKGVKLYATKPLSHPLNDOLRATLGYQOQVFGHS-TNGFDL 664
Db 458 TKNDYGTSVNLGYTPRYTTKQVSLGKGVNFFENDYNSKSDSSNKRTTYGSAVTVLGEFPV 517
QY 665 STRT-----LEH---EISRSITIQNGWNR---YSLRYRLDKLKTQADPETWQDLPVDFV 713
Db 518 NENNSYVYGLGHTYKNKISNFAL---YNRNLTYIQMKFKGNGIKTN-----DFDFS 565
QY 714 NGKPSQELLAGVAVHKTVADNLVNPMPRGYRQRYSLFVGGSGLV-----SDANMALARGI 769
Db 566 FGW-----NYNSLNRGYPTPKGVKASLGGRTVTPGSDNKYKLSADV 607
QY 770 SGVYSGFDNAYGNSRAH---QMTGGIAGYIWSDNENHPYRLRFPAGDQSIGVAHSL 827
Db 608 GGFYPL-----DRDLHVVYSAKASAGYANGFGKRLPFPQTYIAGISLIRGVAISI 660
QY 828 SP-----ISDKGYLTGGOVLAVTAE--YNEFMKD-----LRNAV 862
Db 661 GPMALYEGNGSGTGTFPKKISSD--VIGGNALATASABELIYPTPEVSDKSONTQVTSLEF 718
QY 863 GDIGNAY-----DKGFTNDPKIGAGVGNRNASPVGOVVDVATVVK 902
Db 719 VDASVMMTKMKSDKNGLESNVLKDLPDYGKSSRTIRASTGVGFQWQSPGPPVFFSAKFI 778
QY 903 K-BEGNPFKLHFETGTP 919
Db 779 KKYENDVEQFQFQSIGSE 792

RESULT 8
US-08-433-522A-4
Sequence 4, Application US/08433522A
Patent No. 6013514
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: THOMAS, Wayne
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: SIA, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & McBurney
STREET: 6TH Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,522A
FILING DATE: 12-SEP-1995
CLASSIFICATION: 435


```

Db      349  LRFEGTIVSADSTLNGENRMOQDGTMYNSQVLELGRILDRIFGFEYENRIDIPIGNSDE 408
Oy      565  ---VYOSKKVPLLYEVVADSKPRDGOI--GLMGSDTGTRLVTKF--EHNLI----- 608
Db      409  VDVVYKVK-----ERNTGSINFGIGYTESGISYQASVKODNPLGTGAANSIAG 457
Oy      609  NRBDYQAGAEELRLSE---DKQGVKLYATKPLSHPLNQLRATLTGYQOEVEFGHS--TNGFDL 664
Db      458  TKNDGTSVNLGTYTERPYTTKQGVSLGCVNFPFENYDNKSDTSSNKKRTTYGGSNVTGLGFPV 517
Oy      665  STRT-----LEH---EISRSIIIONGMNRT---YSLRYRLDKTLKTOAPPTMMDPLVDPV 713
Db      518  NENNSYVYGLTGYTKKINFPLE---YNRLNLYIOSMKFKKNGICITN-----DFDPS 565
Oy      714  NCKPQOEALLAGVAVHKTVAIDLNVPMKGYRQRYSLVEYSGSGLV---SPANNAIIRAGI 769
Db      566  FGM-----NNSLNNGYPPFTKGVKASLGGRTTIPGSDNKKYKKSADV 607
Oy      770  SGVYSFGDNVAGNSRAH--QMTGSIQAGIYIMSDNFNHPVRLRFPAGGDSIFGVAHDSL 827
Db      608  QGFYPL-----ORDHLVNVSAKASAGVANGFQNKRLPYQTYTAGGISLGLGFAVGS 660
Oy      828  SP-----ISDKGYLVGOVLAAGAE--VYEEFMKD-----LRLANF 862
Db      661  GPNATVAYEYCGSGTGTFKKIISD--VIGGNALATPASAELIVPPVPSKSONTVPRTSLF 718
Oy      863  GDIGNAY-----DKGFTNDFTIGAGVGVGRMASPVGOVVDVATGV 902
Db      719  VDASVYMTKWKKSDKNGLESVDLKRLPDYCKSSRIASGYGVGMOSPIGLPLVSTAKPI 778
Oy      903  K--EGGNPIKHPFICPFP 919
Db      779  KYENDVEQFQFSIGGSF 797

RESULT 10
US-09-135-166-2
: Sequence 2, Application US/09135166
: Patent No. 6083743
:
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Slim & McBurney
: STREET: 6TH Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/135,166
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/433,522
: FILING DATE: 12-SEP-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
: TELECOMMUNICATION INFORMATION:

```

```

? TELEPHONE: (416) 595-1155
?
? TELEFAX: (416) 595-1163
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 797 amino acids
? TYPE: amino acids
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US-09-135-166-2

```

Query Match	Score	DB 3;	Length
3.68;	171.5;	DB 3;	797;
Best Local Similarity	19.88;	Prod No 10-05;	

Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

QY	263	DVIHHDGEVYIDYRA-----VEYRGGGA--DDKAFTTVADEVPLIGDVFHHGKETK	313
Db	72	DVAMHDEGDVLVSVAAKSIDVAKIKGNSVPTPEALKONIDANGCFKGDVL----	IREK 127
QY	316	KNLIENASAEHGYFDGRWMDRSVDY---LPDNTADSVLIYDTGYQVFEDEVFTIDPK	372
Db	128	LNFEAKSVKEHVAASVGRY-NATVEPIVMTLPNNRAEIIQIINEDKKAASLTF-----	K 181
QY	373	TNOLTDPDKLPVKRBLLEBOLLTVMKGEAY-----NLQAVRLSNDLIATRYENNVT	425
Db	182	GNM-SVSSSTLOEOMELOPDSMMKLMGKFEGBAOFEXKLOISIR-----DYLNNGYA	232
QY	426	EIVFEPEOIODQVSEFSSSSRTEPPQVD-----ESTLVEIETVELTQG	472
Db	233	KAOITFTDQVLNDEKTKVAVITIDVNEGLOYDRSARRIIGNLGGSALEPLLSAHLNDT	292
QY	473	I-LMDISPEFSASLJODKLNLYAAKRHLIDMPD-----DRLVAINHDDGVNRSIL-	524
Db	293	FRRSIDIADIE---NAIKAKLGERGYGATVNSVPEDDANKLTALTYVDAGRRLTVRG	348
QY	525	---GRISDAVSAVARRILPDES---ENEVIDLP-----ERT--ALANKRTPAD-----	564
Db	349	LRFEGNTVASDSTLREQEKROEGTWYNSQVELCKITLDRGFEFEVYENHIDIPINGSND	408
QY	565	---VYQSKKVLPLYFVASDKPRDQOI---GLWGSJDTGTRLVTKF-EHNLI-----	608
Db	409	VDVYVKV-----ERNTGSINRGIGYGESGISYQASVKQDNFLGTGAANVING	457
QY	609	NNDGYQAQAEILRSE---DKRGVLYLTKPLSLHLDQLRATIGYQOEVPFGHS-TNGFDL	664
Db	458	TKNDYGTSVNLGYTEPEYFTKDGVSGLGNNFEVENDNSKSPSTSSNRYRTYGGSVNVLGFPV	517
QY	665	STRT-----LEH---EISRSIIIONGGWNR---YSLRYRLDJKTKQAOPETMQLPVDPV	713
Db	518	NENNSYVYGHLGHTYNNKISNPALE---YNRMLYIOSAKFEKNGIKTN-----DEPDS	565
QY	714	NGKRSQELLAGVAVHKTVDNLVNPGRYRORYSLEVGSSGLV---SDANMAIARAGI	769
Db	566	FGW-----NYSNLIRGIFYPRKGVAKASIGCRVITPGSDNNKYKLSADY	607
QY	770	SGVYSFGDNAGYNSRAH---OMTGGIAGYIMSDNFHNVRYRLRFPFGQOSIRGTASHSL	827
Db	608	QGFYPL-----DRHLMWVASAKASHGANGFNKRLLPFOYTYAAGGSLRGFAYSI	660
QY	828	SP-----ISDKGYLTGGOVYLVAGTAE--YNYEFMKD-----LRLAVF	862
Db	661	GPNAIYAENGSQGTGFFKSISSD-VIGGNMIAATASABLIPTPFVBDKSGQNTVYRISLF	718
QY	863	GDIGNAV-----DGKFTNDRIAGAGVYRMASFPYGOVRYVDATGV	902
Db	719	VDAASVWNTWMSKDKNGLESYVILKRLPDYKSSRINASTYGVGFOWQSPIGLPLVESYAKPI	778
QY	903	K--EGGNPILAHFICTPF	919
Db	779	KXYENDVDEQFQFSIGSF	797

```

; Sequence 4, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-135-166-4

Query Match      3 6%; Score 171.5; DB 3; Length 797;
Best Local Similarity 19.8%; Pred. No. 1e-05;
Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

QY 263 DVIIDHLEPVYIDYRA-----VEVRGECA-DDKAFYTVADDEVPIILGDFHHGKETK 315
DB 72 DYKAHQEGVLVSVYAKLISDVYIKKNSVITFEALKNLDANGKRVGYVL-----IREK 127
QY 316 KULIENASAEHGFEDGRMLDRSDVVI--LPDNTADVSLIYDTGTQYRFDEVVFFETIDPK 372
DB 128 LNEFAKSVKEHVASVGRY-NATVEPIVNTLPNNRAEILQINEDDKAKLASLTF-----K 181
QY 373 TNOULTDPKLPYKRELLDLQLLVNNGEAY-----NIQAVALSNDLIATIRFMVNT 425
DB 182 GNE-SVSSSTLDEQMLQDPDSWMKLMGNKEFGAQFEKDLQSIK-----DYLLNGYA 232
QY 426 EIVPEREQIIONDOVSFEQSSSRTEPAQVD-----ESTLPEVIVTELTG 472
DB 233 KQOITKTDVQNLDEKTKVNTIDVNGLOYDLRSARITIGNLGMSALEPLLSALHLNDI 292
QY 473 I-LMDISPIEFSASNLIDQKLMVAARARLLYDMPD-----DRYLAINHDDGVNRSLI-- 524
DB 293 FRRSDIADVE---NAIKAKLGERGYSATVNSVPEDDANKTLATTLVVDAGRRLTVQ 348
QY 525 ---GRISAVSAVARAILPDES---ENEVIDLP-----ALANKPIPAD----- 564
DB 349 LRFEGNTVASDSTLQKEMQOEGTWNYSQLVELGKIRLDRTGFEFTVENRIDPINGSNDE 408

```

```

QY 565 ---VYQSKVPLVYVVASDKPRDGOI--GLGWSGDTGTRLVTKF-EHNL-I----- 608
DB 409 VDVVYKVK-----ERNTGSINIGIGYTESGISYQASVKQDNFLGTGAAVSIAG 457
QY 609 NRDCYQAGAEIRLSE---DKKGVKLYATKPLSHPLNDQLRATLTIGQVEFGHS--TNGFDL 664
DB 458 TKNDYGTSVNLGYTEPEYFTTKDVSLSGNVFPEENDNSKSDTSNNKRTTYSNVTLGFPV 517
QY 665 STRT-----LEH---EISRTIIONGMNRT---VSLRYRLDKLTQAPPETWQDLPVDFV 713
DB 518 NENNSYVGLGHTYTKNISNFALE---YKRNLYIOSMKFKGICITN-----DFDFS 565
QY 714 NGKPSOALLAGVAVHKTVDNLVNPNGYRQRYSLVEGSSGLV---SDANMAIARAGI 769
DB 566 FGW-----NNSLNGYPTPTGKVASLGGRYTIPESDMKYKRLSADY 607
QY 770 SGVTSFGDNAYGSNRAN--QMTGGIOAGYIWSDNFNHVPYRLRFPAGDOSIRGYANDSL 827
DB 608 QGFYPL-----DRDLHWVYSASAKASAGYANGFGNKRLPFYQTYTAGIGSLRGFAYGSI 660
QY 828 SP-----ISDKGYLTGGQYAVGTAE--YNEEFMKD-----LRLAVF 862
DB 661 GPNAIYAEYNGSGTGTFKKISSD--VIGGNAIATASAEILVPTPEVSDKSQNTVRSLEF 718
QY 863 GDIGNAY-----DKGFTNDTKIGAGVGVWMAVSPGVQVRYDVATGV 902
DB 719 VDAASVNTKWKSKDNKGLSEVTLKRLPDYKSSNRIRASTGVGFQWOSPFIQGLVSYAKPI 778
QY 903 K--EGNPRIKLHFFIGTFP 919
DB 779 KKYENDVEQFOFSIGSF 797

RESULT 12
US-09-135-166-6
; Sequence 6, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155

```

TELEFAX: (416) 595-1163
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 797 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-135-166-6

Query Match 3.6%; Score 171.5; DB 3; Length 797;
 Best Local Similarity 19.8%; Pred. No. 1e-05;
 Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

```

QY 263 DVIHDLGEPVYIDYRA-----VEVRGEGA-DDKAFTTVADEVPLLIGDVFHGKYEK 315
DB 72 DYKAOEGDVLVSVAAKSIISDVKIKGNSVITEALKQNLANGKRVGDVL-----IREK 127
QY 316 KLIENASAEHGYFDGRMLDRSDVY---LPDNTADVSLIYDTGTQYRFEDEVFFETIDPK 372
DB 128 LNEFAKSVKEHYSVGRY-NATVEPIVNTLPNNRAELIIOINEDDKAKLASLTF-----K 181
QY 373 TQGLTDPDKLPVKRRLLEQLTLVNNGEAY-----NLQAVRALSNDLIATRYFNKAVNT 425
DB 182 GNE-SVSSSTLOEQMELQDPSMMKLMGNKFEQAQFEKDLQSTR-----DYLLNNGYA 232
QY 426 EIVFPEREQIONDQVSEFQSSSRTEPAQVD-----ESTLEPIETVELTDG 472
DB 233 KAQIKRTDVOLNDEKTKVNTIDVNEGLOYDLRSARITGNLGMASLEPELISALHNDT 292
QY 473 T-LMDISPIEFSASNLIDQKLNVAAKARHLYDMPD---DRVLAINHDDGVNRSIL--- 524
DB 293 FRRSDIADVE---NAIKAKLGERGYGSATVNSVPDQDANKTLAITLVVDAGRRLTVRQ 348
QY 525 ---GRISDAVASAVARAILPDSES---ENEVIDLP---ERT---ALANRKTTPAD----- 564
DB 349 LRFEGNTVSAADSTLROEMRQOEGTWYNSQLVEIGKIRIDRTGFEFTEVENRIDPIGNSDE 408
QY 565 ---VYQSKKVPRLVVEVASDKPRDQOI--GLWGSDTGTRLYTKF--EHNLI----- 608
DB 409 VDQVYKVK-----ERNTGSIINFGIGTGTESGISTYQASVKODNFTLGGAASVING 457
QY 609 NRDGIQAQAEQLRLSE---DKKGVKLYATKPLSHPLNDQLRATLGYYQEEVFGHS-TNGFDL 664
DB 458 TKNDYGTGVNLTGTEPYFTKDGVSIGLVNFENYDNSKSDTSSNKKRTTYGNSVTLGFPV 517
QY 665 STRT-----LEH---EISRSIIQNGMNR---YSLRRLDLKLTQAPEPTMODLPEVDF 713
DB 518 NENNSYVGLCHTYYKIKISNPALE---YNNRLYIQSMKFKGNGIKTN-----DFDS 565
QY 714 NGKPSQALLAGVAHVAKTVADNLVPMRGYRORYSLEVGSSGLV---SDANMATAIRAGI 769
DB 566 FGM-----NYNSLNGGYPTTKGVKASLSGGRVITPQSDNKKYYLSADV 607
QY 770 SGVYSEFGNVAGSNRAH--QMTGCIQAGYIMSDNFNVHVPYRLRFPAGDOSIRGYAHDLSL 827
DB 608 QGFYPL-----BRDHLVNVSAKASAGYANGFGNKRRLPFQYVTAIGISLRCGAYYSI 660
QY 828 SP-----ISDKGYLGGQVLAVGTAE--YNYFEMD-----LRLAVF 862
DB 661 CPNATVAYCNGSGTGTFKTISSD--VIGGNALIVTASAEILLVPTPFVSDKSQNTVRTSLF 718
QY 863 GDIGNAY-----DKGFTNDTKIGAGYGVIRMAKSPVGVQVRVDVAICV 902
DB 719 VDAASVNTWKMSDKNGLESDYLKRLPDYKSSRIKRASTGAGFQWQSFITGLPVSYAKPI 778
QY 903 K--EEGNPIKLHFTITP 919
DB 779 KKYENDVEQFQSIGSGF 797

```

RESULT 13
 US-08-942-046-2
 ; Sequence 2, Application US/08942046

```

; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Two Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 MIS-Jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-942-046-2

```

Query Match 3.6%; Score 171.5; DB 4; Length 797;
 Best Local Similarity 19.8%; Pred. No. 1e-05;
 Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

```

QY 263 DVIHDLGEPVYIDYRA-----VEVRGEGA-DDKAFTTVADEVPLLIGDVFHGKYEK 315
DB 72 DYKAOEGDVLVSVAAKSIISDVKIKGNSVITEALKQNLANGKRVGDVL-----IREK 127
QY 316 KLIENASAEHGYFDGRMLDRSDVY---LPDNTADVSLIYDTGTQYRFEDEVFFETIDPK 372
DB 128 LNEFAKSVKEHYSVGRY-NATVEPIVNTLPNNRAELIIOINEDDKAKLASLTF-----K 181
QY 373 TQGLTDPDKLPVKRRLLEQLTLVNNGEAY-----NLQAVRALSNDLIATRYFNKAVNT 425
DB 182 GNE-SVSSSTLOEQMELQDPSMMKLMGNKFEQAQFEKDLQSTR-----DYLLNNGYA 232
QY 426 EIVFPEREQIONDQVSEFQSSSRTEPAQVD-----ESTLEPIETVELTDG 472
DB 233 KAQIKRTDVOLNDEKTKVNTIDVNEGLOYDLRSARITGNLGMASLEPELISALHNDT 292
QY 473 T-LMDISPIEFSASNLIDQKLNVAAKARHLYDMPD---DRVLAINHDDGVNRSIL--- 524
DB 293 FRRSDIADVE---NAIKAKLGERGYGSATVNSVPDQDANKTLAITLVVDAGRRLTVRQ 348
QY 525 ---GRISDAVASAVARAILPDSES---ENEVIDLP---ERT---ALANRKTTPAD----- 564
DB 349 LRFEGNTVSAADSTLROEMRQOEGTWYNSQLVEIGKIRIDRTGFEFTEVENRIDPIGNSDE 408

```

```

OY 565 --VOSKRPVLYEVASDRKRDQI--GJGWSGDTGRILVTKR--EHNLL-----608
Db 409 VDVAYTKV-----EKRTSINFGIGYGESGISTOASVYKODNFIETGAANSIAG 457
OY 609 NRDGYOAGAEELRLE---DKKGVLYATKPLSHPINLOJRALTICYOQOEVEGHS--TNGFDL 664
Db 458 TKNDGTSVNLGTEPEYPTFKDGVSELGGVNFEPENDKSKDPTSSNYKRTTYTGSNVTIGFPV 517
OY 665 STRT-----LEH--EISRSIIIONGGNNR---YSLRLDLKLTQCAPETWODLPVDEY 713
Db 518 NENNSYVGLGHTYNNKISNFALE---YNRMLYIYSMKFEKNGIKXTN-----DFDES 565
OY 714 NGKPSOEALLAGVAVHKTVAIDLNPMMRGYKQRYSLFVSGSGLY---SDANMAIRAGI 769
Db 566 FGW-----NYNSLRGRFPTPYGVASASLGRVYITIGSNNKTKYKISADY 607
OY 770 SGVYSFGDNATGNSRAH--OMTGGIOAGYIWSDNFNHVPYRLRFEAGGDOSIRGVAHDSL 827
Db 608 QGFYPL-----DRDHLWVVSASAKASGYANGFENKRLPLPYQYRTVAGGISTLGRFAYGSI 660
OY 828 SP-----ISDKGYLITGGOVAVLGRAE--VYNEPRKD-----LRLAPF 862
Db 661 GPNAIYAEYCGNGSGTGTFPKTISD--YIGGAALITASAEILVPTPEVSDKSQNTVPTSLF 718
OY 863 GDIGNAY-----DKGFTNDTKIGAGVYGRMASPVQGVAVDVAATGV 902
Db 719 VDAASVWNTKWKSPKNGHLESVULKRLPDYCKSSIRASTSVGQWQMSPIGLVFTYAKPI 778
OY 903 K--EENGRPIKHLFFICTPFF 919
Db 779 KKYENDVDEQFQFSIGGSF 797

```

RESULT 14
US-08-942-046-4
Sequence 4, Application US/08942046
Patent No. 6264954
GENERAL INFORMATION:
APPLICANT: CHONG, peje
APPLICANT: THOMAS, Wayne
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: SIA, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & Mcburney
STREET: 6TH Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,046
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/433,522
FILING DATE: 12-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-732 MIS-jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163

```

; INFORMATION FOR SEQ ID NO: 4:
;
;   SEQUENCE CHARACTERISTICS:
;       LENGTH: 797 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-08-942-046-4

```

Query Match 3.6%; Score 171.5; DB 4; Length 797;
Best Local Similarity 19.8%; Pred. No. 1e-05;
Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

QY 263 DVIIHDLGEPPYIDYRA-----VEVHGEGA-DCKAFITVADEVRLIGVPHHGKYEK 315
 Db 72 DYKHAQEDVLYVSVAKSIISDYKIKIGNSVIPLEALKONLDANGFVGVYL-----IREK 127
 QY 316 KNLINASAEHGFDGRWLDKRSVDVI---LPDNTADSVLYIDTGYRFEVEVFETIDPK 372
 Db 128 LNEFAKSVEKHYASVAGR-NATVEPIVNTLPPNRAELIQLINEDKAKLASLNF-----K 181
 QY 373 TNLQITDPDKPYVKRELLEQLLTYNNMGEAY-----NLQAVRLASNDLIATRYFNKVT 425
 Db 182 GNE-SVSSSTQIEOMELIOPDSMKMLCKNKEGAGFEKDLQSIK-----DYLLNNGYA 232
 QY 426 EIVEPEREIOINDOVSEFOSSSSRTPEAODY-----ESTLEPIVETVELTQG 472
 Db 223 KAQITKTDVQUNDKTKYVNTYIDVNEGLQYDLDNSARITGLMGMSALEPLLSALHLNDT 292
 QY 473 I-LMDISPIEFSASNLIODKLNLYAAKARHLYDMPD---DRYLAINHDDGVKRSIL--- 524
 Db 293 FRRSIDIADVE---NAIKAKLGEKGYASATVNSVPDDEDKNTLATIYLVDAGRRLTVRQ 348
 QY 525 -----GRISDANSAAKARILPDES---ENEYIDLK-----ALANKKTPAD----- 564
 Db 349 LREEGNTVVSADSTLRQEMRQOEGTWNYSOLVELGKIRLDTGTGFEETVEYENRDIPIINSNDE 408
 QY 565 ---YEOSKAKVLYVFAVSDKPRDQOI---GLMGSDSTGRVLTKFE-EHNLL----- 608
 Db 409 VDVVYKXK-----ERNGTSGINFIGIGYGETESIGTQASVKODNPLGTGAANVING 457
 QY 609 NRDOYOGAELRLSE---DKKGVKLYATKPLSHPLNDOLRATLGYOQOEFGHS-TNGFDL 664
 Db 458 TKNDYGTGSVMNGYEPETFKDQVSLGANGVFENYDNKSPDTSNRYKRTTGYSNVTLGFPV 517
 QY 665 STRP-----LEH---EISRSIIONGGNKR---YSLRYRLDKLKTQAPETWODLPYDEV 713
 Db 518 NENNSYVVGIGHYTKNSISNFALE---YNRRLYIQSMMKFKNGIKITN-----DEDES 565
 QY 714 NGKPSQEALLAGVAVHTVADNLYNPMRGYRQRYSLLEVSGSGLYV---SPANMAIARAIGI 769
 Db 566 FGM-----NYNSLRKRGIFPPKGVKAKSLGCGVYTLPGSDNKTKYKLSADY 607
 QY 770 SGVYSFGDNAYGSNRAH---OMTGGIOAGYIWSDNFNHVPYRLRFFAGGDOSIRGVAHDSL 827
 Db 608 QGFEYPL-----DRDHLWVWASAKASGANGPNKRRLPEYQYTTAGYIGSLGFAVYSI 660
 QY 828 SP-----ISDKGLTYLGGQVLYAVGTAE---TYEEKMD-----LRLAVF 862
 Db 661 GPNATYAEYNGSGGTGTEKKISSD-VIGGNALATATASAEILVPTPVSDBKSONTVRTSLF 718
 QY 863 GDIGNAY-----DKGFTNDTRIGAGVGRBMSAPVGOVRVNDVATGV 902
 Db 719 VDAASVWNTKMKDKDKNLEDVYLKRLPDTYKSSRIASTGCVGFOWOSPIGPLVSTAKPI 778
 QY 903 K--EGGNPIKLHFFIGTGP 919
 Db 779 KKYENDVDEQFQSGSGSF 797

RESULT 15
US-08-942-046-6
; Sequence 6, Application US/08942046
; Patent No. 6264954


```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 393-2000
TELEFAX: (415) 393-2286
TELEX: 340817 MACPAG SFO
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3031 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-689-008-2

Query Match      3.4%; Score 159.5; DB 1; Length 3031;
Best Local Similarity 19.2%; Pred. No. 0.001;
Matches 201; Conservative 131; Mismatches 360; Indels 357; Gaps 51;

QY 43 PAHDPAI-----NQAKGNPPVLLT-PRQIQARL--NAAGLNARQSOAL- 84
DB 1720 PHSLSAVEYQTMAGVPAQMDQARAGLAGVVASNPQDYRQALFAQALTYNTSTRMGLT 1779
QY 85 ---DVVNFDDQSPI-SRIGESPPGLID-MSVIEETTPLELFAQESTEMGINNDYI 139
DB 1780 RLKDIQSFQSPQAVEAAAAQSTRLSLMYPVPEYQPLMEQWLSA-----HPND-- 1829
QY 140 PEYQGEQPNSEVVVPPLEPERKPGILKRLYARLFNDGVNKKVPRLKAKFYQSSQSGETSAT 199
DB 1830 -----TALREHMLHPGCPDPAKGLARQ-----AGYQQLNAGRLAA 1866
QY 200 GSSHQKTEPYANIKALEITQESANDLNGSIPRLKOTLVAAARAVGYDIDISTIRNST 259
DB 1867 EOSFQ-----SALQINS-----HDAD-----SL 1884
QY 260 GEVDYIIHDLGEPYIDYRAVEVRGEGADDKAFTVADEVPLLIGDFHNGKETKKNI 319
DB 1885 GGMGLVSNRQGTJA-EARRYEEEAADPK-----TADKMRPALAGMAVSGEYASVROLI 1938
QY 320 ENASAEHGTFDGRW-----LDRSDVY----- 341
DB 1939 ---AAHOYTEAKQOLATLARQGYTGATLMLADLQSTGQIAAEQRYGLISREPNN 1994
QY 342 -----LPDNTADV-SLIYDTGYQ--REDEVYFFITDPKTNLTTPDCKLPVK 386
DB 1995 OLALMGLARVDMAQNGTAEARQLLSRVGPQYASQVGEIEVSGIMAAQSOTSDSARKVSI 2054
QY 387 RELLEQL-----LVNMGAEVYNIQ-----AVRAL-----SMD 413
DB 2055 REANAQARDPVRRINTLNALDQOQSDVAEAGRVMOPIIANPVTAODRQALITYTGSNGD 2114
QY 414 LIATRYFMVNTIEIVFPREQIIONDOVSFEQSSSRTEPAQVDESTLEPVIETVELTDCI 473
DB 2115 AMTRQLAGLSPADYSPAIRSIA-EEMEIKODLASRLSVNSPVLIRKALIQDPPTGAR 2173
QY 474 LMDISPIEFSASNLIDKLNLYAAKARHLYDMPDRV-----LAINHDDGVNRSILGR 527
DB 2174 GVAADAVFERQGDVYHAAWMAIRIASTRTIDLSPDORLSYATEYMKISNPVAAAR-LIAPL 2232
QY 528 SDAV-SAVARAIIIPDESE-----NEVIDLPERTALANKPTADVYQSKVPLVYVVASDK 581
DB 2233 GDSGSAISGALLPEQVYTOQLRMGISVAQSDLNOGRDQQAQADHLAPALQADPEATIS 2292
QY 582 PR-----DCQIGLGWSDTGTRLYTKFEHNLINRDQYQAGAEIRLSEDKKGV--KL 630
DB 2293 KFLALRLYNGHGKPKKALEID-----LAVLRHNQDDADQAQAAVQAVNSHNSLARLR 2347
QY 631 YATRPLSLPLNDOLRATLGYQOEYFGHSTNGPDLSTRLLEHISHSITONGGWNRTYSLR 690
DB 2348 AMDGVQSSPMDARAMLAAVADQADGH-----GQRTIE-----DLRRAYDLR 2389
QY 691 YRL-----DKTKTQAPET-----MODLPVDYVNGKPSQQA----- 721
DB 2390 LQOQVEGTRRAASGAAQOEDAL--APFSTNPPRRPGYGHOTELAPVYTGGSYSABAASPD 2446
QY 722 ---LAGVAVH-KTVADNL---VNPNGRYQRYSLSEVGSSGLVSDANMALA-----RAG 768
```

```
DB 2447 TSDQMLSSIAGQIRTLLENLAPSIDGLGFRSR-SGEHG-MGRLTENAPIVGLRPLQAG 2504
QY 769 ISGYVSFGDNAYGNSNRHQMTGTGIQAGYIMSDNFN-----HVPYRLFFAGGDQISRGY- 822
DB 2505 ASA-----LTFESITPTMWSGNLNTSGSYVDV--RY--GTMMGVQAYN 2543
QY 823 AHDSLSPIQSKYLTG--QVLAVGAETVNEYEMKRLRLAVGEDIGNADKGTNDTKI 879
DB 2544 QYDS-----YTNMAGRDOORIAAGTAE-----AGFAPDVQF 2573
QY 880 G-----AGVGVRAWSPYGOVVRVATGVK 903
DB 2574 GNSWVRADVG---ASPIGFPTIVNLGVE 2599

RESULT 17
US-09-268-347-49
; Sequence 49, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: LOOMORE, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-49
```

```
Query Match      2.9%; Score 138; DB 4; Length 2314;
Best Local Similarity 19.3%; Pred. No. 0.042;
Matches 189; Conservative 132; Mismatches 315; Indels 342; Gaps 53;

QY 68 ARLNAAGLNARQSOA-----LDVNFDDQSPISRIGESPPGLDMSVIEETTPIS 119
DB 435 AELQSGGLFTFTPTNASTDKTYVGTDLKFTDMS-----NTALEDTTTRIT 479
QY 120 LDEL-FAQESTEMGINPNDYIPEYQGEQPNSEVVVPPLEPERKPGILKRLYARLFNDGVN 178
DB 480 KDKIGFSNKA-----GTVDENKPYLDK----- 501
QY 179 KYPRLEKAKFYQSSQGET--SAIGSSHQKTEPYAN-IRKALEDTITQESAMDLNGSIPRLR 235
DB 502 --DKLKVGSSTLNGLTYNNNTIGSNKOIOVGADGDKFA-----DVNVNVSMAA 549
QY 236 Q--TALVAAARAVGYDIDLSITIRNSIGEVDYIIHDLGEPYIDYRAVEY-----RGE 285
DB 550 KFGTTRITREEIGFADAD-----GKYD-----KKSPIYDKKQLOVGVKYYTKDSGI 595
QY 286 GADDKAFTVADEVPLLIGDYVFNHGKYEYTKKMLIENASAEHGTFDGRWLDRSDVYLLPN 345
DB 596 NAGDQKISNVKAT-----DDTDAVYTKOLQOVQDA-----DALQSFST--RDE 639
QY 346 TAD---VSLIYDTG--TOYRFDEVF-----FTIDP--KTNQLTID 379
DB 640 KQGEFTISMLYNSGNTPNFTFTTFRAGENGISISNDIAKGVKVGIDPITNGLTTPKLTVG 699
QY 380 PDKLPVKRELEQLLVN-----MGEAVYNIQAVR-----AL 410
DB 700 SDKDGKTQOLVIEQVASNGTKNIIINGVSPITLPSITNAGSVRTTEGNTTTSDEDSKAAS 759
QY 411 SNDLIATRYFMVNTIEIVFPREQIIONDOVSFEQSS-----SSSRTEPAQVDESTLEP 462
DB 760 IGDILINTG-FNLKN-----NSNSYGVFSYTNVDFIDGNATTAKAVYVDETNN-- 804
QY 463 VLETFELFDGIIMDISPIEFSASNLIDKLNLYAAKARHL-----YDMDDDRVL 511
DB 805 --QTSKVTYDVAVDEKTIETLIGDN--GTYNKIGVKTITTTTNNANGATYNSTTDNAL 859
```

QY	512	AHHDDVNSIIIGRISDAVSAVARAIL-----PDESENEVIDLPREXTALANKKPPAD	564
Db	860	VNAKDIAENLNTLAKELHHTTKGTADPALOTFKYKKDGAOTDETTIV-KGOTGONKTVNT	918
QY	565	VYOSKKVPLVYVASDOKPRDDQIGLWMSDGTGLVTKFEHNLINRDG-----Y	613
Db	919	L--KLKGENCLGVATNK--DGTVTFGINQSGLK--AGDSTLLKGLISTIKNPASNEQI	971
QY	614	QAGAE-----LHLSDDKKGV-----KLATAPRLPSHPLNDLQRLALG	649
Db	972	QVGDADGVFAKVDKGNSSGIGDTSITITDQIGFPGANGSLDITPPR--HLTKDKLKVG--	1027
QY	650	YQOEVEFCHSTNGFDLSRTLEHEISRSIION-----GGMNRTYSRLRYLD-KLKTQAPR	702
Db	1028	-EVEI---TMTGINAGGKKTITIQOSDIDITQNSNDAYTGS--RYVDLKTLEBKIKISAA--	1079
QY	703	ETWODLPEVDPEVNGKPSOEALLAGVAVHK--TVAD-----NLVNPGRGYRQRYSEV----	751
Db	1080	-----KTAQNSL-----HEFSVADDEQNHFTVSNPSSYDTSKTSPTVIFA	1120
QY	752	GSSGLVSDAMMAIARAGISSVYSEFGDNAYGNSRAHQDMTGGIQTAGITWSDNPFHHVYRLRF	811
Db	1121	GENSIITTKVKKGVYRGT-----DQTKGLTTPKLTGNGNNKGKIVIDS-----	1163
QY	812	FAGGDOSIRGVANDSLSPISDKCYLGTGGVLANVGALEYEERFMKRLRLAVFGDIGNAYDK	871
Db	1164	-KDCQNTITIGLS--NTLANVYNDG--AGHALSGLGLANDT-----DKTRAASITCDVLANA---	1211
QY	872	GFTMDTKIGAGCYGRNAS	889
Db	1212	GFNLQ---GNGCEAVDFVS	1226

RESULT 18
US-08-621-944A-4
Sequence 4, Application US/08621944A
Patent No. 6440425
GENERAL INFORMATION:
APPLICANT: SASAKI, Ken
APPLICANT: HARKNESS, Robin E.
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Slim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/621,944A
FILING DATE: 26-MAR-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/478,370
FILING DATE: 07-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-587
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163

```

; INFORMATION FOR SEQ ID NO: 4
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 1833 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;
US-08-621-944A-4

```

Query Match	2.9%;	Score 136;	DB 4;	Length 1833;
-------------	-------	------------	-------	--------------

Matches 191; Conservative 128; Mismatches 351; Indels 298; Gaps 49;

QY	22	PLMTSOLAQONNPANIINIHPADTAI--NQAAAGBPVLLPEQIOA---RLNA-AGLN	76
Db	294	PYLKQOLKQSGVAITITDNGIDGNKKISMLAGSSHANDAVTIEQLAKPPLNAGAGIS	353
QY	77	AKPOSALDVNPFDOQSPISRIGESOPPLGD-----MSVIEETPLSL--BELFAOEST	129
Db	354	VTPEIISVDKSGVWTPATYNIQKVTTELMSDGTSDKFSVSGSGTNLSLVTAHEHLASYLN	413
QY	130	EMGINPND-----YIPEYQGEQBPSEVYVPTLEPEKPGILKLYARLNDGVNKKPRLK	184
Db	414	EVRNTPADSALOSFTVKKEEDDDAALITVADTKR-----NAGAVSTILKL	458
QY	185	AK-----FYOSSQSGTSAIGSSHQKTEPYANIKAALEDITQESA-----M	225
Db	459	GKNGLTVAATKKDGTVTGTGLSDQSGLT--IKSTLNDNG-LTVADTMEQI-QVGANGIKFT	514
QY	226	DLNSIP--RLRQJALVAARAQVGYIIDLSITNSIGEVDVYIHDGEPYIYIDRAYEV--	282
Db	515	NVNSGNSPQGTIANPARRITRDKIGFAGSDAV-----DNPK-YLDQDOLQG	560
QY	283	-----RGEGADOKAFPTVADDEVPL-----IGDFPHG-K	311
Db	561	NVLTNTNGINAGGALITGLSPILPSTADQSSRIEIGNTIQDKDKSNASINDLNTGN	620
QY	312	YETKNLLENNAABHGTFDGRWILDRSYDI--LPDTPADVSLIYDTGTQYREDEVF--FT	368
Db	621	LKNNNPIDEVS-----YDIYDFANGNATATVYHDPA--KTSKVYDVN	665
QY	369	IDPKNOL--TTPDKLPVKRELLEQLITVNMEAVNLAVALRSLNDLATGR-FMNVYT	425
Db	666	VDOTTHLGTDDNKKRKGVTTKLN-----KISANGNATNPNNSDSE	709
QY	426	EIVPEREQION--DOVSFEQSSSRTEPAQVDESTLEPIYETVELDTGLMDISPESA	484
Db	710	DALVNADIDIEMLNTLAKELHTTKGTADTAILOFTFKKDDENNADANAIVTG--OKNA	767
QY	485	SNLIQDKLNLVAAKARLYMPDDR-----VLAINHDGVNRSILGRISDAVSABARILP	540
Db	768	NN-----QVNTLITLKGENGLIKTDKNGTYFGINTTSGK-----AGKSTLN	810
QY	541	DESENEVIDLPERTALANRRTPADVYQSKKVPPLYEVASDKPRDQIGLGWGSDTGRLY	600
Db	811	DGQ---LSIKNPGFSQIOVGADGVKFAKV-----NNGVYGAG--IDGTRRI	853
QY	601	TKFPHNLINDDGYQAGAEHLRLSEDKKGVYLAKPLSLPHLNOQLRATLTGYQOVBGH--ST	659
Db	854	TRDEIGTTGNG-----SLDK-----SKP-----HLSK	876
QY	660	NGEPLSRTLEHEHSRIION-----GGMNRTYSLRYRLDKLKTQAPPETMODLPVDFV	713
Db	877	DGIAAGKKTITNIOGSEILAQNSHDVATGG--KLYDKTELE-----	915
QY	714	NGKRSQALLAGVAVHK--TYAD-----NLVNPMBGTRORTSLEV-----GSSGLVSDAMM	762
Db	916	NKLSSTAKTQMSLHEFVSADQGNFTYNSPYSYDTSKTSIDVITFAGENGITTKYKK	974
QY	763	AIARAGISGYVSGDNAYGNSRAHOMTGTGIOAGYIMSDNFNVHVPRLRPFAGGDSIGCY	822
Db	975	GVYARVGI-----DQTKGLITPKLTVGNNGKGLYIDSQN-----GQNTITGL	1016
QY	823	AHDSLPSIDSKGYLTGGQVYLAAGTAELNYEFMKD--LRLAVFGDI-----GNAY	869

```

Db      1017 SNTLANVTNKG-----SVRTTEOG-NIIKEDKTRASIVDVSAGFNLOGNGEAV 1067
OY      870 DKGFNDT 877
Db      1068 DFVSTYDT 1075

RESULT 19
US-08-945-567D-4
; Sequence 4, Application US/08945567D
; Patent No. 6448386
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF
; TITLE OF INVENTION: MORAXELLA
; FILE REFERENCE: 1038-745 MIS
; CURRENT APPLICATION NUMBER: US/08/945,567D
; PRIOR FILING DATE: 1996-04-29
; PRIOR APPLICATION NUMBER: 08/431,718
; PRIOR FILING DATE: 1995-05-01
; PRIOR APPLICATION NUMBER: 08/478,370
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/621,944
; PRIOR FILING DATE: 1996-03-26
; PRIOR APPLICATION NUMBER: PCT/CA96/00264
; PRIOR FILING DATE: 1996-04-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1833
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-08-945-567D-4

Query Match      2.9%; Score 136; DB 4; Length 1833;
Best Local Similarity 19.7%; Pred. No. 0.042; Indels 298; Gaps 49;
Matches 191; Conservative 128; Mismatches 351;

```

```

Db      666 VDDTTHLTGTDNKKLGKVTTKLN-----KTSANGNTATNFVNSSDE 709
OY      426 EIVPEREQION-DOVSPEOSSSSTPEAOYDESTLEPIETVELTDCIIMDISIESA 484
Db      710 DALYNAKDIAENLNTLAKIEHTTKGTADTALOTFTVKRVNDENMADNATTVG--QKNA 767
OY      485 SNLIQDKLNLVAARARHLHYDMPDR-----VLAINHDDGVNSIIGRISDAVSAAVARILP 540
Db      768 NN-----QVNTLTILGGENLNKTDKNGVTGINTTSLK-----AGKSTLN 810
OY      541 DESENEVIDLPERTALANRKTTPADYOSKVPPLYVFAVSDKPRDQIGLGMSDPTGRV 600
Db      811 DGG-----LSIKNPTGSEQIQVADGVKPAV-----NNNGVYAG--IDGTTRI 853
OY      601 TKFEHNLINRGYQAGAEIRLSEDKKVKLYATKPLSHPLDQLRAITGYQOEVPFGH--ST 659
Db      854 TRDEIGFTGNG-----SLDK-----SKP-----HLK 876
OY      660 NGFDLSTRTELEHISRTION-----GNNRTYSLRYRLDKLTQAPPTWQDLPVDFV 713
Db      877 DGIAGGKKITINIOSGELAONSHAVTGG--KITDLKTELE----- 915
OY      714 NGKPSQALLAGVAHK--TVAD-----NLVNPARGYRQARYSLEY-----GSSGLVSDAM 762
Db      916 -NKISSPAKTAQNSLHFSVADDEQNNFTVNSPYSDTSYSTPITPAGENGITTKYK 974
OY      763 AIAAGISGYVSFEDNMGSRRAHOMTGIGAGIYSDNFHVPYRLRFFAGDQOSIRGY 822
Db      975 GVAVRGI-----DQTKGLTTPRLTYGNNNGKVIDSON-----GNTITLGL 1016
OY      823 AHDSLSPISDKGYLTGGOVLAFTAENYEMKD---LRLAVFGDI-----GNAY 869
Db      1017 SNTLANVTNKG-----SVRTTEOG-NIIKEDKTRASIVDVSAGFNLOGNGEAV 1067
OY      870 DKGFNDT 877
Db      1068 DFVSTYDT 1075

RESULT 20
US-08-621-944A-3
; Sequence 3, Application US/08621944A
; Patent No. 6440425
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/621,944A
; FILING DATE: 26-MAR-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/478,370
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I

```

```

1 REGISTRATION NUMBER: 24,973
2 REFERENCE/DOCKET NUMBER: 1036-587
3 TELECOMMUNICATION INFORMATION:
4 TELEPHONE: (416) 595-1155
5 TELEFAX: (416) 595-1163
6 INFORMATION FOR SEQ ID NO: 3:
7 SEQUENCE CHARACTERISTICS:
8 LENGTH: 1992 amino acids
9 TYPE: amino acid
10 STRANDEDNESS: single
11 TOPOLOGY: linear
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019

```

Query Match	2.98; Score 136; DB 4; Length 1992;
-------------	-------------------------------------

QY	22	PLMTSOLAQONNPNANIINHVAHOTAI-NQAKAGNPVLLTPEIDOA---RLN-AGIN	76
Db	453	PLYLOKKOLKVGSAITIDGIDAGNKKISLNKAGSSANDAVITIEQKAKAPPLNAGAGIS	512
QY	77	AKPOQALDVNFFDOOSEPIRIGEOSSPPLGLD-----MSVIEETPLSL-EELEAOEST	129
Db	513	VTPTEISVDASGNAVTAPTYNIGVATTELNSOGTSODKRSVKSGSNNLSLVTAHEHLASYLN	572
QY	130	EMGINPND-----YIPEYOGEOQNPSEVVYPPPLLEPEKPEGLIKRLYARLEPNDGVNKKVPRK	184
Db	573	EYNRTADASALOSFYKAEEDDDANAITVAKDPTK-----NAAVASILKLK	617
QY	185	AK-----FQSSQSGEATMAGSSHQKTEPEYANIKKALEDITOESA-----M	225
Db	618	GKNGLTAVTKKDDGYTFFGLSDOSGLT--IGKSTLNNDG-LTYKDTNEOI-QYGGANGIKFT	673
QY	226	DUNGSI-P-RLKOTALVARAVGYVDLDLSIIRNSIGEVDIITHDGEVYIDYRAVEY-	282
Db	674	NYVNGSNPFGIANTARITRDKIGFAGSDGAV-----DTNKP-YLDQOKLOY	719
QY	283	REGADDAKFYTVADEVPLL-----IDGVFHHG-K	311
Db	720	NVKTNTNTINMGKAKITGSLPTLPIADOSSNIELGNTIODKDSMAASINDILNTGTGN	779
QY	312	YETKKNLIENASAEHGFYDGRWLDRSDVYI-LPNDTADVSLIYDTGTQYREDEYVF-FT	368
Db	780	LKNNNNPIDFVS-----TYDLYDRANCNATATYTHDTAN--KRSKYVYDVN	824
QY	369	IDPKTNOL--TTDPKLPYKRELLBOLLTVNMGAYNLQAVALSNDLIATRY-FNMVNT	425
Db	835	VDDTTHLGTDDNKKLCVKTTKLN-----KTSANGNATATFNNSSD	868
QY	426	EIVFEREDION-DOVSFEQSSSRTEPAQVDESTLEPIETVELLDGILMISPIEFA	484
Db	869	DALVAKKIDAEUNTLAKEIHTTKGTADTALOTFTYVKKYDENNMADNANITVG--QKKA	926
QY	485	SNLIDOKDLNVAAKARHLHYMDPRD---VLAINHDDGVNRSILGRISDAVASAVARAIIIP	540
Db	927	NN---QVNTLLTKGENGINKTKDRNGVTFGINNTSGK-----ACKSTLN	969
QY	541	DESENEVIDLPEPTALANRKTTPADYIOSSKKVPLYFVASDKRDCQIGICMGOSDPTGRLV	600
Db	970	DGG-----LSIKNPTSEQIOVGADGVKFAKV-----NNNGCVVAG--IDGTRRI	1012
QY	601	TKFEHNLLNPDQYQAGAELELSEDKGVKLYATKPLSHPLNDQLBATLIGYQGEVFGH-ST	659
Db	1013	TRDELGFPGTNC-----SLDK-----SKP-----HLSK	1035
QY	660	NGFDLSRTLEHEHSRSITON-----GGWNTYSLRYRLDKLKTQAPPETWQDLPVDEV	713
Db	1036	DGINAGCKKIIINIGCEIAONSHDAVTGG--KIYDLKTELE-----	1074
QY	714	NGRPSQELLALGAVAHK-TVAD-----NLVPMNGYRORYSLEY-----GSSGLVSDAMN	762
Db	1075	NKJSSJTKTKWNSJHBSVADGEOGNNFTVSNPSSYSDTSKTSVDITFAGGNGIITTKVKR	1133

[illegible]

RESULT 21
US-08-945-567D-3
; Sequence 3, Application US/08945567D

	Query Match	2.9%	Score 136;	DB 4;	Length 1992;
	Best Local Similarity	19.7%	Fred. No. 0.045;		
	Matches 191;	Conservative 128;	Mismatches 351;	Indels 296;	Gaps 49.
Oy	12	PLMTSOALAOQNPNPANIHHVPAHPTAI--NQAKAGNPVLLTPEQIOA--RLNA-AGLN	76		
		: : : : : : : : : : : : : : : : : : : : : : : :			
Db	453	PYLDDKQLKVGSAVLTIDNGIDAGKKISLAKGSSANAAVITIEQAKAKPPLNMGAGIS	512		
Oy	77	AKPQSOALDVNPFDDOSPISRIEGOSPPLGID---MSVIEETTPLSI--ELLEAQEST	129		
		: : : : : : : : : : : : : : : : : : : : : : : :			
Db	513	VPTPEISVDAKGSAVTAFTYINIGVTTTELSNGTSIDKSFVKSGSTNNSLVTAEBHLASYLN	572		
Oy	130	EMGINPND-----YIPEYQGEQNPSENVVYPPLEPERPBLIRLARLENDGVNKKVPKL	184		
		: : : : : : : : : : : : : : : : : : : : : : : :			
Db	573	EVNRTADASALQSFIVKEEDDDANAIYVAKDTK-----NAQAVSTILKL	617		
Oy	185	AK-----FYOSSQSGEISAIGSSHOKTEPYANIKAALEDITOESA-----M	225		
		: : : : : : : : : : : : : : : : : : : : : : : :			
Db	618	KGNGLTVAATKKQGYTFGLSQDSGLT--IGKSTLNNDS--LYAKDINQEI--QVGANGINKFT	673		
Oy	226	DINGSIP--RLKOTALVAARAVGYVDIDISITRNSIGEVYVITHDLGEPVYIDYRAVEY-	282		
		: : : : : : : : : : : : : : : : : : : : : : : :			
Db	674	NVNGSNPFGIAINTARITRDKIGFAGSDGAV-----DTNKP--YLDQDLQVQ	719		
Oy	283	-----RGGADDKRAFTYADEVPLL-----IGDVFFHG-K	311		
		: : : : : : : : : : : : : : : : : : : : : : : :			
Db	720	NVKTITNTGIMAGKAITGSLPTPLTADSOSSKNIELGNTIQDKDKSNASINDILNTGFN	779		
Oy	312	YETKKNLIEMNSAEHGVEFGRLIDRSVDY--LPDNTADVSLTYDGTGYRDEVEVF-ET	368		

```
Db      780 LKNNNNIDFVS-----TYDIDVFANGNATTAATVTHDTAN--KTSKVYVDVN 824
Qy      369 IDPKTNOL--TTDPDKLPYKRELLEQLLTVMGSAVNLQAVRALSNDLIATRY-FNNVNT 425
Db      825 VDDTTHLTGTDDKKKLGVTTKLN-----KTSANGNTATNFVNSSDE 868
Qy      426 EIVPEREOLION-DQVSEFQSSSRTEPAQVDESTLEPVIEVTELDTGLMDISPIEFA 484
Db      869 DALVNAKDIAENLNTLAKELHTTKGTADTAQLQFTYKVKDENNADANMTTVC--QKNA 926
Qy      485 SNLIQDKLNLVAAKARLLYMPDR-----VLAIHDDGVNRSILGRISDAVSAVARAILP 540
Db      927 NN-----QVNTLLTKGEGNLIKTDKNGTVTFGINTTSGLK-----AGKSTLN 969
Qy      541 DESENEVIDPERTALANRKTADVYOSKRVPLVFAVASKPRDGOIGLWGSPTGTRLY 600
Db      970 DGG-----LSIKNPTGSGOIGVADGVKFAKY-----NNGGVGAG--IDGTTTRI 1012
Qy      601 TKFEHNLINRDGYQAGAELELSEDKKGVKLATKPLSHPLNDQLRATLGYQOEVEFGH-ST 659
Db      1013 TRDEIGFTGTNG-----SLDK-----SKP-----HLSK 1035
Qy      660 NGSDLSRTLEHEISRSIION-----GGMNRTYSLARYLDKLKTQAPPETWODLPVDFV 713
Db      1036 DGINAGKKTITNIQSGEIAQNSHDAVTG--KIYDLKTELE-----1074
Qy      714 NGKPSOALLAGVAVHK-TYAAD-----NLVPMRGYRORYSLEV-----GSSGLVSDANM 762
Db      1075 -NKISSAKTAQNSLHEFSVADEQNNFTVSNPYSDTSTKTSVITFAGENGITTVYVK 1133
Qy      763 A1ARAGISGVSEFGDNAVGSNRAHQMTGGIOAGYIMSDNFHNYRRLRFFAGDOSIRGY 822
Db      1134 GAVRVGI-----DQTKGLTTPKLVGNNGKGIIVDSQN-----GQNTITGL 1175
Qy      823 AHDSLPISDKGLTGGQVLAAGTAENYEFMKD--LRLAVFGDI-----GNAY 869
Db      1176 SNTLAVNTNKG-----SVRTTEGG-NIKDEKTRAASIVDVLSAGFNLOGNEAV 1226
Qy      870 DKGFNTD 877
Db      1227 DFVSTYDT 1234

RESULT 22
US-09-268-347-48
; Sequence 48, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Locomotion, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-8660
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 2048
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-48
```

Query Match 2.9%; Score 136; DB 4; Length 2048;

Best Local Similarity 19.6%; Pred. No. 0 051;

Matches 188; Conservative 133; Mismatches 356; Indels 284; Gaps 48;

```
Qy      22 PLMTSQALQOQNNPANIINHVAHDTAI-NOAKAGNPVLLTPEDQIA--RLNA-AGLN 76
Db      508 PYLDKQLVGVSAITIDNGIDAGNKKISYLAKGSSANDAVTEQLKAAPLTINAGAGIS 567
Qy      77 AKPQSGALDVVNDDQSPISRTGQSPPLGLD-----MSVTEETPLSL--EELFAQEST 129
Db      568 VPTPEISVDKAGSNVATPYNTNIGVKTTELSNDSGTSKFSVKGSGTNNSLVTAHLASYN 627
```

```
Qy      130 EMGINPD-----YIPEYQEQPNSEVVPPTEPEKPKGLIKRLYARLFNDGVKKVPLK 184
Db      628 EVNRTASALQSFYVKEDEDDANAIYAKDTTK-----NAGANSIILKLK 672
Qy      185 AK-----FYQSSQSGEYSAIGSSHOKTEPYANIKALDEITQESA-----M 225
Db      673 GKNGLVATKTKDGTVFTEGLSDQSGLT--IGKSTLNNDG-LTVKDTNEDQ-QVANGAIKFT 728
Qy      226 DLNCSIP--RLRQALVAARAVGYTIDLSIIRNSIGEVYIINDLEPVYIDYRAEV- 282
Db      729 NVNGSNPQGTGIANTARTTRKIGFAGSDGAV-----DTNKP-VLDDDKLQVG 774
Qy      283 -----RGEADDAKFTTVADEVPL-----IGDFVHHG-K 311
Db      775 NVKTTNGINAGGAKAITGLSPTLESIADQSSRNIELGNTIODKKSNAASINDLINLNGFN 834
Qy      312 YETKKNLIENASAEHGYFDGRMIDRSVDYI--LPDNTADVSLIVDTGTQYRFDEVF-FT 368
Db      835 LKNNNNIDFVS-----TYDIDVFANGNATTAATVTHDTAN--KTSKVYVDVN 879
Qy      369 IDPKTNOL--TTDPDKLPYKRELLEQLLTVMGSAVNLQAVRALSNDLIATRY-FNNVNT 425
Db      880 VDDTTHLTGTDDKKKLGVTTKLN-----KTSANGNTATNFVNSSDE 923
Qy      426 EIVPEREOLION-DQVSEFQSSSRTEPAQVDESTLEPVIEVTELDTGLMDISPIEFA 484
Db      924 DALVNAKDIAENLNTLAKELHTTKGTADTAQLQFTYKVKDENNADANMTTVC--QKNA 981
Qy      485 SNLIQDKLNLVAAKARLLYMPDR-----VLAIHDDGVNRSILGRISDAVSAVARAILP 540
Db      982 NN-----QVNTLLTKGEGNLIKTDKNGTVTFGINTTSGLK-----AGKSTLN 1024
Qy      541 DESENEVIDPERTALANRKTADVYOSKRVPLVFAVASKPRDGOIGLWGSPTGTRLY 600
Db      1025 DGG-----LSIKNPTGSGOIGVADGVKFAKY-----NNGGVGAG--IDGTTTRI 1067
Qy      601 TKFEHNLINRDGYQAGAELELSEDKKGVKLATKPLSHPLNDQLRATLGYQOEVEFGHSTN 660
Db      1068 TRDEIGFTGTNGSLDKSKPRLSKD--GINA-GGKKTINIOSGETQA-----N 1111
Qy      661 GFDSLSTRLEHEISRSIIONGGMNRTYSLRYLDKLKTQAPPETWODLPVDFVNGKPSOE 720
Db      1112 SHDAVT-----GG--KIYDLKTELE-----NKISSST 1135
Qy      721 ALLAGVAVHK-TYAAD-----NLVPMRGYRORYSLEV-----GSSGLVSDANMAIARAGI 769
Db      1136 AKTAQNSLHEFSVADEQNNFTVSNPYSDTSTKTSVITFAGENGITTVYKNGVAVGI 1195
Qy      770 SGVYSEFGDNAVGSNRAHQMTGGIOAGYIMSDNFHNYRRLRFFAGDOSIRGYAHDLSLP 829
Db      1196 -----DQTKGLTTPKLVGNNGKGIIVDSQN-----GQNTITGLSNTLAVN 1237
Qy      830 ISDKGYLTGGQVLAAGTAENYEFMKD--LRLAVFGDI-----GNAYDKGFNTD 876
Db      1238 TNDKG-----SVRTTEGG-NIKDEKTRAASIVDVLSAGFNLOGNEAVDFVSTYD 1288
Qy      877 T 877
Db      1289 T 1289
```

RESULT 23

US-09-134-001C-3159

; Sequence 3159, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

;; PRIOR APPLICATION NUMBER: US 60/064,964
;; PRIOR FILING DATE: 1997-11-08
;; PRIOR APPLICATION NUMBER: US 60/055,779
;; PRIOR FILING DATE: 1997-08-14
;; NUMBER OF SEQ ID NOS: 5674
;; SEQ ID NO 3159
;; LENGTH: 10182
;; TYPE: PRT
;; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 2.9%; Score 135; DB 4; Length 10182;
Best Local Similarity 20.0%; Pred. No. 0.93;
Matches 166; Conservative 106; Mismatches 293; Indels 266; Gaps 38;

QY 27 QALAOQNPNP-TINHVPADTAIQAAGNPVLLTPBQIQAARLNAAGLNAKPOSQLD 85
DB 2695 QAEBQAIQHANSVINNGATSSQINDAK--NTVEQAQRDYEAQSMLRADKSQLOS-AVD 2751
QY 86 VVNP-----DOSP-----ISRIEESPPLGLDMSYEETPLSLEELFAQESTENG 132
DB 2732 TLNRDVLTKNDKRPASVRRYNEAISIRKELDTAKADASTLRNTSPSYQVADALNKINT 2811
QY 133 INP--NDYIPEYGEQPNSEVVPPLEBEKPLIKRLVRLFNDGVNKPRLKA----- 185
DB 2812 VQPKVQAIALLOPKENNELVQAK-----KRL-----QDAVNDIPQTQHTQOT 2856
QY 186 -KFTYSSGSETSAIGSSHQKTEPYANIKALEDITQESAMDLNGSIPRLROT--ALVA 242
DB 2857 INNVDKQREAFERALTASQORVID--NGDATQETISEKS-----KVEQAQALNTNA 2905
QY 243 RAVGYDDIDLSIRNSIGEVDIHDLGEPVYIDYRAVEVRGEGADKRAFTVADEVPLL 302
DB 2906 K-----SLUR-----YETKKNLIENASAEHGYFDGRWLD-----ADKNELOTAYNK---L 2923
QY 303 IGDVFNHCK-----YETKKNLIENASAEHGYFDGRWLD-----RSVDVIL 342
DB 2924 IENVSTNGKRPASIRQYETAKARIQNOINDAKNEARILGNDNPQVSQVTOALNKIKAIQ 2983
QY 343 PDNTADVLSLID-----TGQYRFDEVYFTTIDPKTNQLTTD-----PDKLVKREL 389
DB 2984 PKLTETAINMLQNKENTELVNAKNLENAVNDT--DP--THGMTQETINNNAKREAOQNEI 3041
QY 390 LEOLLTVNMGCA-----YNLOAVRALSNDL-----LATREYFN--VNFRI 427
DB 3042 QKANMIINMGDATAODISSEKSYEQVLOALQNAKNDLRADKRELQTAIVNKLITQNVNTNG 3101
QY 428 VFPE-----REQIONQVSEFQSSSSRTPEPAQVDESTLEPIYETVELTGDILMDISP 479
DB 3102 KKPSSIONYKSAARRIEN-----QYNTAKNEAHNVLENT--NPVTNAVEDALRKINAIQ 3154
QY 480 IEFASNLIOCK-----LNVAAKARHLVMPDDR 509
DB 3155 EYTAIKNILQCKENSELVRRAKEKLQAIINSOPSLNGMTOESINNYTKRREAOAIASSA 3214
QY 510 VLAINHDGVRSTIG--RISDAVSAAVARA---ILPDES--ENEVIDLPERTALANKT 561
DB 3215 DTIINNGASIEQITTEKIRBEATNALNEAKHULTADTSLKTEVRKL--SRGCDTNNK 3273
QY 562 PADVYQSKKPLVYFVASDKPRDQIGLGWSDTGTRLVTFEHNILNRDGYOAGAEURL 621
DB 3274 PSSV-----SAYNNTIHSLOSSETQ 3293
QY 622 SEDKKGKLVATKPLS-----HPLNDQIRATLGYQOEYFGHSTNGFDLSTRTL 669
DB 3294 TENNANTTI--NKPIRSVEEYNNALHEVNOQLNORLTDITINLOPL--ANKESLKEARRNL 3349
QY 670 EHESTRSIOGNGWNRITYSLRYRLDKLQAPPETWODLPVDFVNGKRSQE 720
DB 3350 ESKINET-VOTDGMTQOSVENTKQAKIKAOSSIAQTLL--INNGASDQ 3396
RESULT 24

US-08-961-083-118
; Sequence 118, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 118:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1040 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-118
Query Match 2.8%; Score 134; DB 4; Length 1040;
Best Local Similarity 19.4%; Pred. No. 0.024;
Matches 190; Conservative 124; Mismatches 318; Indels 348; Gaps 53;
QY 24 MTSQALAOQNNPANI--INHVP-----AHDTAIQAAGNPVLLTPBQIQAARLNA 73
DB 52 LTSQLSAVNSQSLSGVGHLEPEPLKIEGYOTIGYIKTKKQONTLSRT---VDGKYSQ 108
QY 74 GLNAKPOQALDVNFD-----OSPISRIEQSPPGLDMSVLEETPLSLEELFAQ 127
DB 109 RDSQPNSTKTSVYHSADELMNMGOGKVSLOGEASGDGL-----SEKSIADNLSSND 163
QY 128 STEMGINNDYIPEYGEQPNSEVVPPLEBEKPLIKRLVRLFNDGVNKPRLKAF 187
DB 144 SFASQVEON--PDKHGS-----VVRPTV--PEQGNPASA----- 194
QY 188 YOSQSGETSAIGSSHQKTEPYANIKALEDI--TOESAMDLNGSIPRLRQALVARAVG 246
DB 195 -TTVQSAEEVYLATNDP-----YKLPETKGTQEPGHE----- 229
QY 247 YYDIDLSTIRNSIGEVDIHDLGEPVYIDYRAVEVRG--EGADKRAFTVADEVPLLGD 305
DB 230 -----GSA-AVREDL--PVYT--KPLETKGTQPGHGEAAVAREEPATYEP 271
QY 306 VF-----HHGKETKKNLIE-----NAAEHIYEPGRMLDSVYIILDPNDTA 348
DB 272 LATKGTQEPGHEKATVBEETLETEPYATKGTQEPHE--EG---ERVBEELP----- 320
QY 349 VSLIYDTGYRFEDEVYFTTIDPKTNQLTTDPRKLPVRELLLEQLLTVNMGAAYMLQAVR 408
DB 321 -ALEVTTNKRTEIONI-----PYTTEIQTDPPLTKNKRRIERO-----GQA----- 360

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 1326
TYPE: PRT
ORGANISM: Acetobacter xylinum
FEATURE:
OTHER INFORMATION: n at positions 15741 and 15767 may be a, g, t, or
US-09-147-236-5

Query Match 2.8%; Score 130.5; DB 4; Length 1326;
Best Local Similarity 19.6%; Pred. No. 0.072;
Matches 197; Conservative 139; Mismatches 386; Indels 283; Gaps 52;

OY 47 TAINOAKGNPPVLT-PEQIOARL---MAAGLAKPOSQAL---DYVNDQDSPT-SR 97
DB 181 TOMPOARAGLGIYVANSQNYRAQLAFAQALTYNTSTREMEGLTRUKDQSFQSQAPVEAA 240
OY 98 IGEOSPLGLD-MSVTEETPLSELEFAQSTEMGINDPIPEYQGPSEVVPPT 156
DB 241 AATOSYROTLSMLPVPNPPTQPLMEQWLSA-----HPND-----AALREHMLHPG 285
OY 157 LEPEKPLIKRLYARLFENDGVNKYPRKA--KFYOS--SGETSAG-----SSHQ-- 204
DB 286 GPPOKAGLARAQAGYQQLNAG-----RLSAAEQSFQALQINSHDADSLQGMGLVMROGD 340
OY 205 --KTEPVANIKALEDTQE-----SAMDNGSIFRLROTALVA-----A 242
DB 341 TAAHPYEEAAMADPKYADRMKRALAGMAVSGDYAAVRQ--LIAHOYTEAKOKLATLA 398
OY 243 RAYGVY--DIDLSTIRNSIGEVYIIDLGPYIDYRAVEVR-----GEGADDK 290
DB 399 ROPQOYTATLMLADLQSTGQVAAEQ-----EYRGILSTREPNNQALMGLARVDM 450
OY 291 APTVADEVPL-----IGDVFHKGKETKKNLENSAHEGYFDGWLDRSDV 340
DB 451 AAGMTAEARQLISRVSPQYASQVGEIEVSG-----LMAAASQ----- 487
OY 341 ILPDTAAVSLIDTQGYRDEVEVFTIDPKTNQTLTDPKLRYKRELEQLLVNKG 400
DB 488 -TSQARKVSTIREMAAPRDPWYRNL--ANALQOOGDAERGRVA--OPTLANPYT 541
OY 401 AVNLQA---VRALSNDLIATRYFMNVTEIFYPEREQIQNDQVSEFQSSSRTEPAQVD 456
DB 542 AODRQAGLITYGSGNDAMTROLLAGLSPADSPAIRSLA--EEMELKODLASRLMVSMP 600
OY 457 ESTLEPVYETVELDGIILDISPIEFSASNLIOQKLNVAAKARHLMDPDDRV----- 510
DB 601 VPLIREALSDPDPGARVAVADLFROGRGDVHARMALRIASTRTIDLSPDQRLSYATEY 660
OY 511 LAINDOCVNSIIGRISDAV--SAVARAILPDESE-----NEVIDLPERTALANKTPAD 564
DB 661 MKISNPVAAR-LAPPLDGTGSGAGNMLPEQOMOTLQLRNGISVAOSDLNKGDDAQ 719
OY 565 VYQSKVPLVYFVAVSDKPR-----DGOIGLQWGSDDTGRVLTKEHNLINDGYOA 615
DB 720 AYDHLAPALQADPEATSPKLLARLYNGHKPKKALEID-----LAVLRHNPDDIDARQA 774
OY 616 GAELRLSDKGV--KYATKPLSHPLDQLRATIGYQOEYVGHSTNGFDLSTRLENEI 673
DB 775 AVQAAVNSNHNLSLATRLMDGVQESPMARAWLMAVADQADGH-----GORTIE-- 824
OY 674 SRSIIIONGWMRTYSLARYLKL-----KTQAPPET-----WQDLP 709
DB 825 -----DLRRAYDL--RLQOVEGTAAASGPGVGAHEEALAPPSINPQSRGYGHQVELG 874
OY 710 VDFYNGKPSQA-----LLAGVA--VHRTVADNL---VNPMBGYRQRYSLVEGSSGL 756
DB 875 APYVGYSYSAEASPDSTDOMLSTIAQOIH--TLRENLAISPIDGIGFGRSR--SGEHG-MGR 931
OY 757 VSDANMALA-----RAGISGYVSPGDNAVGSRAHOMTGIGIAGIYSDNEN-----HVP 806
DB 932 LTEANIPYICRLPIQAGASA-----LTFSTITPTMIWGSQMLNTGVSVDVP 975

OY 807 YRLRFAGDQSIKGYANDSLSPIDKGYLTG--QVLAVGTAENYEFMKDLRLAVFGD 864
DB 976 -RYGTFMATQAAHQAGHSSCGGLD---FLSANHTQRIAGAAGE----- 1015
OY 865 IGNAVYDKGFTNDTKIGAGVGVNR-----ASPVQGVQVAVATGVK 903
DB 1016 -----AGFAPDVQFGNS-----WVRADVCASPIGPIITNVLAGVE 1050

RESULT 27
US-08-323-477-2
Sequence 2, Application us/08323477
Patent No. 6086896

GENERAL INFORMATION:
APPLICANT: Sparling, P. Frederick
TITLE OF INVENTION: ANTIGENIC IRON REPRESSIBLE PROTEINS FROM
TITLE OF INVENTION: N. MENINGITIDIS RELATED TO THE HEMOLYSIN FAMILY OF TOXINS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Imclone Systems Incorporated
STREET: 180 Varlick Street
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10014
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,477
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/920,963
FILING DATE: 28-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Feil, Irving N.
REGISTRATION NUMBER: 28,601
REFERENCE/DOCKET NUMBER: SPA-2-2P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-645-1405
TELEX: 212-645-2054
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1115 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-323-477-2

Query Match 2.7%; Score 129; DB 3; Length 1115;
Best Local Similarity 19.9%; Pred. No. 0.072;
Matches 205; Conservative 102; Mismatches 335; Indels 388; Gaps 45;

OY 56 NPPVLTPEQIOARLNAAGLAKPOSQALDVVNFDDQSPISIRGEQSPPLGLDMSVTEET 115
DB 155 NSRMKSTSVTLQHELWYLGFEKTL-----DYGN-----ERLGE-----SIMNIDPF 194
OY 116 TPLSELEFAQSTEMGINDPIPEYQGEQPNSEVVPPTLEPEKPG----- 163
DB 195 TPKSLANPFA-----DPTYYSNVLEVSFITYSLVDDANPMKWSGEDIYIGRGISEWG 246
OY 164 -LIKRLYARLFENDGVNKYPRLAKEYQ---SSQGETSAIGSSHOKTEPVANIKALAD 218
DB 247 ELLEKWKYKDF-----LPLEKEWQDFPKFEDWMLPEFPEMARWMLKIDPKRSKGYHYVD 300
OY 219 ITQESAMDNGSIPRLROTALVAAR--ANGYDIDLSIIRNSIGEVY-----YIIHGLSP 272
DB 301 ---PLALDIDGD-----GIETVAAGKGFAGALPDRNOGIRTAIGWVSAADDGLVRLNGN 352


```
QY 273 VTIDRAVEYRGADDKAFTVADEVPLLIGVFHHGKYEKKNLIENASAEHGYEDGR 332
      |||
      |||
      |||
Db 353 GLID-----NGAE-----LFGD-----NTR--LADGSFAKHGYAALA 382
QY 333 WLDRAVDLLPONTADVSL-----LYDTGTQRFDEVFETID----- 370
      |||
      |||
      |||
Db 383 ELDSNGDNII--NAADAAPOTLRVWODLONDGISQANELFTLELGISDLAYKDVKN 440
QY 371 -----PKTNQLTTD-----PDKLPVKRELLBOLLVNMGEAY 402
      |||
      |||
      |||
Db 441 LGNGNTLAOOGSTTKTDGTTAKMGDILLADNLSRKKDVELTAEQAKANLAGIRLR 500
QY 403 NLQAVRALNDLIATRYFMNVNTEIIVPEREQIQNDQVSEQSSSSKTEPAQVDESTLEP 462
      |||
      |||
      |||
Db 501 DLREAAALSGDLA-----NML-----KAVSAETKEAQL--ALLDN 534
QY 463 VIETVELTGLIMDISPIESA-----SNLIQDKLNLVAA--KARHLYDM 505
      |||
      |||
      |||
Db 535 LHKMAETDSNMGKSPMRSLSTDMQTQANEGIALTPSQVQLKKNALVSLSDAKAIDA 594
QY 506 PDDRVLAINHDDGVNRSILGRISDAVASAVARAILPDESENEVIDLPERTALANKTPADY 565
      |||
      |||
      |||
Db 595 ADRIVAVLAYTSGDSSSTLYTMS-----EDALNIVKKTNDT 631
QY 566 YOSKRVPLY--VFVADKPRDQIGLGWSDYGT--RLVTKFEH-----NLINR 610
      |||
      |||
      |||
Db 632 YDLAKNIYONLFTRLQPYLNQISFKMENDEFTLDFSGLVQAFNHNKVTNPQAKAVDL 691
QY 611 DGVQAGAEIR-----LSEDK----- 625
      |||
      |||
      |||
Db 692 AEMLAYGELSMYEGRRMLADVVEAKKAGFEDYOKVLQETVALLAKTSGTQADDILQ 751
QY 626 -----KGVKLYATKPR--LSHPLNDOLRATLGYQOEVEGHS----- 658
      |||
      |||
      |||
Db 752 NNGFGINKNVSLTGANDNTLIGAGNDYLEGSGSDTYFEGKFGQDVIYNVATGRK 811
QY 659 -----TNGF--DLSTRTLE--HEISR-----SIIONGWNRTYSRLRLDKLKT 698
      |||
      |||
      |||
Db 812 DLIRFDGITADMLFTTREGNHLLIKAKDSSGOVTVOSYQNDG-----SCAIRIDE--- 863
QY 699 QAPREWMODLPVDFVNGKSPQSEALLAGVAVHKTVDNLVPMRGYKQORSLEVGS--SGL 756
      |||
      |||
      |||
Db 864 -----IHFDNGKYLDAV-----TVKKELYOQSTGDSRLRYAYQSGSTLNGG 903
QY 757 VSDANMAIARAGISGYVSTGDNAVGSNRHQMGTGQIAGTMSDNNHVRRLRFRAGCD 816
      |||
      |||
      |||
Db 904 LQD-----DLYYGADGNDLLGDAANDSIYSGN-----GND 934
QY 817 OSIRGAHDSLSPISDKGYLTGGQVAVGTAEYNYEFMKDLRLAVFGDIGNAV--DKGFTN 875
      |||
      |||
      |||
Db 935 TLDDGGNDALVGYNGNDALNGE-----GNDHLNGEDGND--TLIGAGANDYLEGSGS 987
QY 876 DTKI--GAGVG 884
      |||
      |||
      |||
Db 988 DTYVFEGBFG 997

RESULT 28
US-09-134-001C-4547
; Sequence 4547, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
```

```
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4547
; LENGTH: 675
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (18)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-134-001C-4547
```

```
Query Match 2.7%; Score 128.5; DB 4; Length 675;
Best Local Similarity 18.6%; Pred. No. 0.034;
Matches 152; Conservative 96; Mismatches 267; Indels 303; Gaps 43;
```

```
QY 187 FYOSOSGETSAIGSSHQKPEPANTAKALEDITQESAMLNGSIPTRLQTLVAARAVG 246
      |||
      |||
      |||
Db 29 FYASKDKEINNTIDALEDK-----NEKQVYKSSYSKSD--NGEVE-----MTERPIK 75
QY 247 YPIDLSIINNSIGENVDIHDLGEPPYIDYRAVEYRGADDAKFTVADEVPLLIGDV 306
      |||
      |||
      |||
Db 76 IY-----NSLGVADINIQDR-----KIKVSKKKKRYD----- 103
QY 307 FHHGKYEYTK--NLLENASAEHGYEDGRW--LDRSDVILLPDMTADVSLIYDTGTQYRFD 362
      |||
      |||
      |||
Db 104 ---AQYKIKTNVGNIDRNVQFNFVKEEDGMKLDMDHVSIIIPGMQDOSI----- 149
QY 363 EYVFETIDPPTNQLTDPDKLPYKRELLBOLLVNMGEAVNLQAV-----RALSD 413
      |||
      |||
      |||
Db 150 -----HIENLKSERGI--LDRNVE--LANGTAVEIGIVPKNVSKKDYKALAKE 196
QY 414 LIATRYFNWNTIELVPEREQIQNDQVSEQSSSPTEPAQVDESTLEPIETVELTDLGI 473
      |||
      |||
      |||
Db 197 L-----SISEDYIKQOMQON-----WVQDDTFRP--LKVYKKHDEY 230
QY 474 LMDIS-----PIEFSASNLI--QDKLNLVAKAKRHLYDMPDRVLA----- 512
      |||
      |||
      |||
Db 231 LSDFAKFFHLTTNETESRNPPLGKATSHLGGYGPRINSEELKQKEYGYGDDAVIGKGL 290
QY 513 -----INHDDGVNSILGRISDAVASAVARAILPDESENEVIDLPERTALANKTPADY 565
      |||
      |||
      |||
Db 291 EKLDRKKLQHEHDYRVYI---VDNSNTIAHTLI---EKKKKDGKDIQTLIDAKYQKSI 343
QY 566 YOSKRVPLYVFVADKPRDQIGLGWSDGTRL-----VTKREHNLINRDGYQAG 616
      |||
      |||
      |||
Db 344 YNNMK-----NDTGSCTAIHPQTGELLALVSTPSYDVYPPMYGMSNE--YN-- 388
QY 617 AEIRLSEDKKGVRLVATKPLSHPLND--QLRATLGYOQOEVEFGHSTNGFSLTRLEHISR 675
      |||
      |||
      |||
Db 389 ---KLTEDEK-----EPL---LNKFQITTSFGSTQKIL---TAMIGLNKKTLDDKTSY 432
QY 676 SIIONGWNRTYSRLRYLQKLTQADPEWODLPV--DFVNGKPSQSEALLAGVAVHKTIV 732
      |||
      |||
      |||
Db 433 K-IDGKGWQK-----DK-----SWGGYNVTRYEVVNGNIDLKQAIES----- 468
QY 733 ADNLVPMRGYRQORSLEVSS-----GLVSDA-----NMAIAAGISGYSEF 776
      |||
      |||
      |||
Db 469 SDNI-----EFARVALEDSKKFEKGMKLLGVGEDIPSDYPPYMAOISKNLNDNETLLA 522
QY 777 DNAVGSNR-----AHQMTGQIAGYIWSDNFNHV-----PYRLRFFAGD 816
      |||
      |||
      |||
Db 523 DSGYGGGELLINVOILSIYSALENNGNINARILLADTKKVKKKINISENINLTDGM 582
QY 817 QST-----RGYAH-----DLSLSPISDKGYLTGGQV-----LAV----- 844
      |||
      |||
      |||
Db 583 OQVYVNTKHEIDIRSYANLIGKSGTAELMKKQGETEQIQMFTSYDKDNPMMAALINVKD 642
QY 845 ---GTAEYNYEFMKDLRLAVFGDIGNAVYDKGTNTQK 878
      |||
      |||
      |||
Db 643 VODKGMASTNAKIS-----GKVDLEYENGK 669

RESULT 29
```

```

US-08-962-859A-2
: Sequence 2, Application US/08962859A
: Patent No. 6127345
: GENERAL INFORMATION:
: APPLICANT: Burnham, Martin K. R.
: TITLE OF INVENTION: No. 6127345el Glucose 6-Phosphate
: TITLE OF INVENTION: Dehydrogenase Gene
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dechert, Price & Rhoads
: STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2793
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compulide
: OPERATING SYSTEM: Windows
: SOFTWARE: FASTSEQ for Windows Version 2.0b
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/962,859A
: FILING DATE: 03-NOV-1997
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/035,072
: FILING DATE: 21-JAN-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Jackson, Arthur E
: REGISTRATION NUMBER: 34,354
: REFERENCE/DOCKET NUMBER: GM50001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 609/520-3254
: TELEFAX: 609/520-3259
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 495 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-962-859A-2

Query Match      2.7%; Score 127.5; DB 3; Length 495;
Best Local Similarity 19.9%; Pred. No. 0.024;
Matches 115; Conservative 93; Mismatches 202; Indels 169; Gaps 32;

OY 148 NSEVVVPTLEPEKPLIKR-LYARLFNDGVNKKVPRKAKFYOSSQSGETSAIGSSHQKT 206
DB 2 SSKVIV--TIFGASGLDAKKRLY-----PSLLDLKXSGNLSKHFAVIGTARR-- 46
OY 207 EPPANIKALEDITOESAMDLNGSIPRLKOTALVAARAVGYIIDLSIIRNSICEVDYII 266
DB 47 -PWS--KEYEVSYYVESILDLDASTEOAOF--SHEYYQ-----S 82
OY 267 HDGGE-PVYIDYRAVEY---RGEADDAKFTVADEVPLLIGDVFPHGKETKNLIEN 321
DB 83 HDVNDLEHTALAKLOELNEKYOAENKLF--FLSMAPQFFGTIAHAKSE---NIV-- 135
OY 322 ASAEHGVEFDGRWLDRAVDVILPDMTADVSLYDTGTQYREDEVVFTIDPKTNOLTTDPD 381
DB 136 -----DCKGFERLL-VEKPFCT-----DYATASKLMDLLAFDE----- 169
OY 382 KLEPVKRELLFOLLTVN--MGEAVNLQAVRALSNDLITRTFNNMNTLVFPEREQIONDO 439
DB 170 -----BOIFRIDHYLG-----KEMIOSIFAVREFANLLFENVW--NKEFINDR 210
OY 440 VSEFOSSSSRTEPAQVDES-TLEPVIE--TVELFDGLIMDISPIEFSASNLIDOKLNLVA 496
DB 211 ITTAEERIGVGERGGYDQSCALROMVGNHTLIQLISLAMD-KPASFTKDEIRAERKIYF- 268
OY 497 AKARHLVMDPDRLVALINHDGCVNRSTLGRISDAVSAVARAILDESENEVIDLPERTAL 556

```

```

DB 269 ---KNLYH-PTDEELKEHFRGOYRS--GKI-DGMKTIYSRSEPN----- 306
OY 557 ANRKTADYQSKKVPPLYEVAVSDKPRDQIGLGWSDGTGLVTKFEH-NLJNRDYO 615
DB 307 VNPESTTEFTSGA-----FVYDSDFR-----GVPFRTGRKRLTEKGTNNIV----- 351
OY 616 GAELRLSEDKKGVKLYATKPLSHPLNDQLRATLGYOQEVGCHSTNGFDLSRTLEHLSR 675
DB 352 -----FKQMSIFGEPLAPNI---LTIYIQPEGFSLSLNGKQVEEF 391
OY 676 SIIONGMNRTYSLRYRLDKLTQAPETMODLPVDFVN 714
DB 392 NLAPN-----SLDYRTDATATGASPEPYEKLITVDVN 423

RESULT 30
US-09-134-001C-5584
: Sequence 5584, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 5584
: LENGTH: 801
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-5584

Query Match      2.6%; Score 125; DB 4; Length 801;
Best Local Similarity 18.1%; Pred. No. 0.09;
Matches 127; Conservative 115; Mismatches 239; Indels 222; Gaps 29;

OY 119 SLEELFAQESTP-----MGINPNDYIPEYQEQPNSVVVPTLEPEKPLIKRL 168
DB 15 SLEEMIKQPYEPMSVSDFODALGINSADSFDR-----LIKILVLEQSLTE-- 62
OY 169 YARLFNDGVNKKVPRKAKFYOSSQGETSA---IGSSHQKTEPPANIKALEDITOESAM 225
DB 63 -----RTRDRYQRQSNKNTSKLIGTISQNKKGFAFLR-----PEDEM 103
OY 226 DLNGSIPRLKOTALVAARAVGYIIDLSI--IRNSICEVDYIIDLGPVYIDYRAV-EV 282
DB 104 DDITFPTKIRAL-----DGDYIVLQKSRGKHKKIE--GVYKSIENHSYQV 152
OY 283 RGEADDAKFTVADEVPLLIGDVF---HNGKETKNLI-----ENASAEHGY 328
DB 153 VGTYSSEAKHFGFVLPDDKRIMQDIFIRKGMILGAVDGHKVLVQITKYADSTDNPEGHNSA 212
OY 329 FDGWRWLDRAVDVILPDMTADVSLYDTGTQYRFEDEVVFTIDPKTNOLTTDPDKLVKRE 388
DB 213 ILGHRNDPQVDIL-----SIYQHGIELEFPDVLQAEAEVPDV--EPESEIGRRD 262
OY 389 LLEQL-LTVNMGAEVNLQAVRALSNDLITATRYFNNMNTLVFPEREQIONDOVSEGOSS 447
DB 263 LRDELITIIDQADAKDL-----DDAIVAKKLKNGNTELF-----VSINDVSYVWEG 309
OY 448 SRTEPAQVDESLPEVETV-----E 468
DB 310 SALDKRAVATSYIVDVRVIMPIRHLNSGICSLNPEEDRLTISCRMEINERGEVYKHE 369
OY 469 LTDDIL-----MDISPIEFSASNLIDOKLNLVAKARHL 502
DB 370 IFDSVISHSNMYTDAVANKIITDDQSEIRSOYKDLTPMLDLAODLSNRLIMRRKRGEID 429

```

```

QY 503 YDMPEDDRLATLNHDGQVRSILTGIRSDAVSAVARAILPDSESENVIDLPERITALANKTP 562
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 430 FDIINAKVLV--NDBGITEVLAR-----EREGG--RLIESFMAIANETV 471

QY 563 ADVYOSKKVPLVYFVASDPRDGOI-----GL---GWGSDGTSLRVTKEFHNLIN 609
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 472 AEHNFKLTVVP-FIRYVHNPQRSDRLRQEFDETNGFMINGTGEDIHPTLLQSIQEEVEG 530

QY 563 ADVYOSKKVPLVYFVASDPRDGOI-----GL---GWGSDGTSLRVTKEFHNLIN 609
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 472 AEHNFKLTVVP-FIRYVHNPQRSDRLRQEFDETNGFMINGTGEDIHPTLLQSIQEEVEG 530

QY 610 RDGVQAGAEILRSDKKGVKILAKRPLSHPLNDLRLATLTGQQEVEFQSHSTNGF---DLS 665
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 531 RPEDQWITSTMILRSMQQA--HYDDVNLIGH-----FGLSAEYUHNHPSPIRRYPDLT 579

QY 666 TRLEHEISRSIIIONGWMNRTSLRYRLDKLKTQAPETWODL 708
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 580 VHRLL--IRKYLIENTS-----MDKKEIHNHMEETLPEL 608

```

```

: RESULT 31
: US-08-537-002A-3
: Sequence 3, Application US/08537002A
: Patent No. 5773282
: GENERAL INFORMATION:
: APPLICANT: TSUSAKI, Keiji
: APPLICANT: KUBOTA, Michio
: APPLICANT: SUGIMOTO, Toshiyuki
: TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
: TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NETMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/537,002A
: FILING DATE: 29-SEP-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 260984/1994
: FILING DATE: 01-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP No. 5773282 yet received
: FILING DATE: 08-SEP-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: TSUSAKI-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: TELEX: 248633
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 963 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: IS-08-537-002A-3

```

```

Query Match          2.68;  Score 125;  DB 1;  Length 963;
Best Local Similarity 24.38;  Pred. No. 0.12;
Matches 98;  Conservative 46;  Mismatches 130;  Indels 130;  Gaps 23;

4  PV-LEANSFMPVALAATVPLMTSOALQOQNPNANIINHVPAHDTAINQA----- 52
|||:::|||||:::|||||

```

```

Db      507 FVELSQQPPFVEGRXYRLTGPHGFAIFALKPVEAVLHLPLSDMAEPAPAEADLPRVH 566
QY      53 KAGNPVLLTEBQIOAR-----LNA-----AGLNKKPOSQA-LDVNFDDOSPISRI 98
Db      567 MGGVEVLVTJLVHERERRELLNALAQTLKESMLAKPQKVALDLDAFR----- 617
QY      99 GEOSPGLDMSVIE--ETTPLELSELEFAOEST-----EMGINPDY- 138
Db      618 -QKDPPLVLTLLQLENNHTLOVSLPLMSFORRGPGLPARTHQPGYFYELESDPEGYR 676
QY      139 -----IPE-----YQGEQNSEVWVPPTEPEKPGILKR-----LYARLFND- 176
Db      677 LLARLKGEFGEGRSUFRAYRGRHRGP--VPEANDLLRPGLAAGEGVWQGLVQDDGDL 733
QY      177 ---VKNVPLK-----AKFYSSQSGETSAIGSSHQKTEPYANIKAALEDITQESAM 225
Db      734 RTERVLPLRLDLPWVLREBGGJLFMERGASRYVALTGLSPGRP-QDLEFVALLEVLSELP 792
QY      226 DLNLSIP-----RLRQT-ALVAAAVGYDIDLSIINSGENVII--HDL----- 269
Db      793 RLRGHAPCTPCLLPGLALHETAVL--RLLG--VRLALLHRALGEVGVGGHPLGLRGL 847
QY      270 -----GEPVYIDYRAVENEGEGADDKAFTTVVDEPPLTIGDV 306
Db      848 GAFLELEGE-VYIVALGAERKG-----TYBEDLARLAYDV 881

```

RESULT 32
 US-08-863-010-3
 : Sequence 3, Application US/08863010
 : Patent No. 6087146
 : GENERAL INFORMATION:
 : APPLICANT: TSUSAKI, Keiji
 : APPLICANT: KUBOTA, Michio
 : APPLICANT: SUGIMOTO, Toshiyuki
 : TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
 : TITLE OF INVENTION: CONVERTING MALTOSIDE INTO TREHALOSE
 : NUMBER OF SEQUENCES: 17
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: BROWDY AND NEIMARK
 : STREET: 419 Seventh Street, N.W., Suite 300
 : City: Washington
 : STATE: D.C.
 : COUNTRY: USA
 : ZIP: 20004
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/863.010
 : FILING DATE:
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/537,002
 : FILING DATE: 29-SEP-1995
 : APPLICATION NUMBER: JP 260984/1994
 : FILING DATE: 01-OCT-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP No. 6087146 yet received
 : FILING DATE: 08-SEP-1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: BROWDY, Roger L.
 : REGISTRATION NUMBER: 25,618
 : REFERENCE/DOCKET NUMBER: TSUSAKI-1
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 202-628-5197
 : TELEFAX: 202-737-3528
 : TELEX: 248633
 : INFORMATION FOR SEO ID NO: 3:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 963 amino acids

```

:
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-863-010-3

Query Match
Best Local Similarity 2.6%; Score 125; DB 3; Length 963;
Matches 98; Conservative 46; Mismatches 130; Indels 130; Gaps 23;

OY 4 PV-LFANRSFMPVALAAYLPLMTSQAALAOQNNPANIINHVPADTAIQA-----52
DB 507 PVLEFSQGFPPVEGKRYRLTGLPHGFALFKVEAVLHLPSPDMAEERPAPEADLPVH 566
OY 53 KAGNPVLLTPRQIQAR-----LNA-----AGLNARQSOA-LDVYVFDQSPISRI 98
DB 567 MGGEPVLLVDTLVHGRGRELNLALQTLKESWLAKPKQVALLDALRF-----617
OY 99 GQSPPLGLDMSVIE--ETTPLSLEELFAQEST-----ENGINPNDY-138
DB 618 -QKDPRLVTLTLQLENHRTLOYSLPLMSPOKREGGLFARTHGQPGYTELSDGFGYR 676
OY 139 -----IPE-----YQGEQPNSEVVVPTLEPEKPGILYK---LYARLFNDG---176
DB 677 LLARLKEGFEGRSRLRAYVRCGRHPR---VPRAVDLLRGLAAGEVWQGLVODGGLD 733
OY 177 --VNKVPRLK-----AKFYQSSQSGETSAIGSSHOKTEPYANIKALEDTQESAM 225
DB 734 RTERVLPRLDLPWVLRPREGGLFEMRGASRRVLTGSLPPGRP-ODLFAALEVRLLESIP 792
OY 226 DLNGSIP-----RLROT-ALVAAAVGYVDIDLSIIRNSIGEVDTI--HDL-----269
DB 793 RLRGHAPGTGGLLPALHETALV--RLLG---VRLALLRLALGEVEGVGSHPLGRGL 847
OY 270 -----GEPVYIDYRAVEVRGEGADKFTTVADVPPLIGDV 306
DB 848 GAFLELGE-VYLVALGAEKRG-----TVEEDLARLAYDV 881

RESULT 33
US-09-024-429-3
: Sequence 3, Application US/09024429
: Patent No. 6165768
: GENERAL INFORMATION:
: APPLICANT: TSUSAKI, Keiji
: APPLICANT: KUBOTA, Michio
: APPLICANT: SUCIMOTO, Toshiyuki
: TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
: TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/024,429
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/537,002
: FILING DATE: 29-SEP-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 260984/1994
: FILING DATE: 01-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 255829/1995
```

```

:
: FILING DATE: 08-SEP-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: YUN, Allen C.
: REGISTRATION NUMBER: 37,971
: REFERENCE/DOCKET NUMBER: TSUSAKI-1B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 963 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-09-024-429-3

Query Match
Best Local Similarity 2.6%; Score 125; DB 4; Length 963;
Matches 98; Conservative 46; Mismatches 130; Indels 130; Gaps 23;

OY 4 PV-LFANRSFMPVALAAYLPLMTSQAALAOQNNPANIINHVPADTAIQA-----52
DB 507 PVLEFSQGFPPVEGKRYRLTGLPHGFALFKVEAVLHLPSPDMAEERPAPEADLPVH 566
OY 53 KAGNPVLLTPRQIQAR-----LNA-----AGLNARQSOA-LDVYVFDQSPISRI 98
DB 567 MGGEPVLLVDTLVHGRGRELNLALQTLKESWLAKPKQVALLDALRF-----617
OY 99 GQSPPLGLDMSVIE--ETTPLSLEELFAQEST-----ENGINPNDY-138
DB 618 -QKDPRLVTLTLQLENHRTLOYSLPLMSPOKREGGLFARTHGQPGYTELSDGFGYR 676
OY 139 -----IPE-----YQGEQPNSEVVVPTLEPEKPGILYK---LYARLFNDG---176
DB 677 LLARLKEGFEGRSRLRAYVRCGRHPR---VPRAVDLLRGLAAGEVWQGLVODGGLD 733
OY 177 --VNKVPRLK-----AKFYQSSQSGETSAIGSSHOKTEPYANIKALEDTQESAM 225
DB 734 RTERVLPRLDLPWVLRPREGGLFEMRGASRRVLTGSLPPGRP-ODLFAALEVRLLESIP 792
OY 226 DLNGSIP-----RLROT-ALVAAAVGYVDIDLSIIRNSIGEVDTI--HDL-----269
DB 793 RLRGHAPGTGGLLPALHETALV--RLLG---VRLALLRLALGEVEGVGSHPLGRGL 847
OY 270 -----GEPVYIDYRAVEVRGEGADKFTTVADVPPLIGDV 306
DB 848 GAFLELGE-VYLVALGAEKRG-----TVEEDLARLAYDV 881

RESULT 34
US-09-134-001C-3172
: Sequence 3172, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3172
: LENGTH: 563
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-3172
```

Query Match 2.6%; Score 124; DB 4; Length 563;
Best Local Similarity 20.8%; Pred. No. 0.06; Mismatches 126; Conservative 103; Matches 195; Indels 182; Gaps 35;

259 IGEVYIINDLGPVYIDYRAVEKRGADKATTVAADEVPLIG-----DYFHHGKYE 313
18 IDBLDINFSD-GLTV-----MSGTSGKSI--IIDAIGOLIGRASSDYVRHGE-- 64
314 TKNLIENASAEHGYPDGRMLDRSVVILPDNTADVSLIDTGTQYRPEVVF-----TI 369
65 -KRAIIE-----GIFD-----IDESKDAI---NILE-SLAIDVDEDFLLVREIFFSGKSI 110
370 DPKTNLTTPDKLPVARELLEQLTVN-----MGEAVNLQAVALSNDLATRYFN 421
111 CRINNOIVLTQD---LRKVMQELLDHGHETQSLKQKHQD---LDDVADNDQYSD 162
422 MVTETVIFPEREIQNDQVFEQSSSRTEPAQY---DESTLEPV---IETVELTJGIL 474
163 LTN-----QYQLSYNOYKKNRKELELESADQALQRLDLMKRLQLEELTEASL 210
475 -----MDISPIEFSASNLIQDKLNLVAAKARHLVDMDDRYLAINHDDGVNRSILG 525
211 KEGEVDLSESDIKRIONS-----EKLIN-----LALNN----- 237
526 RISDAVSAVARATILPDESENEVIDLPERT-ALAN-RKTPADVYOSKVPVPL-----YFV 577
238 -----AHQVLTDSA-----IPRLYELSNLYOTINDIPEKEFVRLEKEDINOQYVL 283
578 ASDKPRDQIGLGWSDYGTRLVTKFPHNLINRDYQAGAE-----RLSEDKGKVL 631
284 LEAKHHEIYDEMA-----NTEFDEQVLINE--YESRMLNLNLRKRGKIDTELIAV 332
632 ATPPLSHPLNDLRATLGVOQEVFGHSTNGEDLSTRTLEH---ISRSIIIONGWNRTYS 688
333 OSK-----LANEIDKIKENYEQ-----STSQLREIKITLYNEVIDIGKLSQ---ERRRV 378
689 LKRLDLKLTQAPETWQDLPVDFVNGKPSQEQALLAGVAAHKTADVNLVNMRCGRORYS 748
379 ARLRHIVSEIOLNOKMDANLE-ISKPLDEPTIEGI---EVEFLISNRGEPKSL 433
749 LEVSSGLVADANKAI-----ARAGISGYVSGDNAYGSMRAHOMTGGIO--AGYIMSD 800
434 NKIASGELSRLMALAKSIVKSGQTALLFDEVDSGVSGAOKMAEKMDIAQYIOVI 493
QY 801 NENHVP 806
DB 494 CISHLP 499

RESULT 35
US-08-072-610-2
; Sequence 2, Application US/08072610
; Patent No. 553133
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,610
; FILING DATE: 19930602

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Aoda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/07686
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELETYPE: 236687
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: PWM3.3.1
US-08-072-610-2

Query Match 2.6%; Score 124; DB 1; Length 1018;
Best Local Similarity 18.6%; Pred. No. 0.16; Mismatches 227; Indels 216; Gaps 31;

30 AQON-NPAN-----IINHPADFTAINQAKGN-----PPVLLTP 63
230 AQENVQPADSNDALFGSLDKDILFDHDKDFEPLQYIAGTAKHYGQELPMKRPVLPV 289
64 EOIARLMAAGLNAKPOSO-ALDVVNEDDQSPISRIGQSPPLGLMSVIEETTPLSLE 122
290 AEPQAQVAEELDATPEDDFALDVT-----ESP-----BEVELVDEE 327
123 LFADESTMGINPDYIPYIYGOEOPNSEVYVPTLEPKPKILRYARLNDG--VAKV 180
328 ATEESTEVG--PTE-----EGPTEELDAP-----EDGRIRRN 360
181 PRLAKFYOSSQSETSAGSSHOKT-EPYANIKALEDITQESAMDLNGSIPRLROTAL 239
361 CRRNRNRVVEETFEAAEGVSEETPGELELATRPD---DPAUD-GTILEETEAT 416
240 VAAVAGYDIDLSIRNSIGEVDVLIHDLGE-----PYIIDYRAVEGEGADKAF 292
417 GEETVEGEETVE-----GEETVEGEAAEGEBELATRPEDDFQLBEPGEGE 463
293 TTVADVEPPLIGVFNHGHKYEKKNLLENASAEHGYPDGRMLDRSVVILFDN----- 345
464 -----GEDEGEDEGEALVAVPYAAE-----PVEVVTAPQPYKPMVA 501
346 -TADVSLIYD-TGYQYRDEVVFFIDPKTNQTLTDPD-----KLVPVRELLEQLLTVM 398
502 PTADETLFVLDLNDLTYADITTSF--EPLFKOILKDPAGAVVYPSKEAVYQVPVAVGP 559
399 GEAVNLQAVRALSNDLATRYFNNVNT-----EIVF-----PEREQIONDVS 441
560 AQEVPTEELMQLED-----DELEGTAEAPEGELVLEGGEPTPEERPGEPTEGEPV 614
442 FEQSSSRTEPAQYDESTLEPIYETV-----ELDGIILMDIS-----PIESASMLI 488
615 EEELEATPEDEFELEPTEGEEVETVGEETAEGEEVEEVAPEVEEVAPEVEEVEV 674
489 QDKLNLVAAKARHLVDMDDRYLAINHDDGVNRSILGRISDAVSAVARAI--LPDESEN- 545
675 PEEVEEVAPEVEEVEEVEE-----VEEVPPEEVEEVEEVEEVEEVEV 719
546 ---EVIDLP-----ERTALNRKTPADVYOSK-----VPLVYFVASDKPR 583
720 EEVEEVEEVAPEVEEVAPEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 779
QY 584 DGQIGLG 590

Db 780 KDELGSG 786

RESULT 36
US-08-719-822B-2
Sequence 2, Application US/08719822B
Patent No. 5874527
GENERAL INFORMATION:
APPLICANT: Barnwell, John
TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby and Darby
STREET: 805 Third Ave.
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-7513
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,822B
FILING DATE: 09/30/96
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 5986/17686052
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)527-7700
TELEFAX: (212)753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1018 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
ORIGINAL SOURCE:
ORGANISM: Plasmodium vivax
IMMEDIATE SOURCE:
CLONE: PVMB3.3.1
US-08-719-822B-2

Query Match 2.6%; Score 124; DB 2; Length 1018;
Best Local Similarity 18.6%; Pred. No. 0.16;
Matches 124; Conservative 100; Mismatches 227; Indels 216; Gaps 31:

OY 30 AQON-NPAN-----IINVPADDTLINAAGN-----PPVLLTP 63
Db 230 AQENVOPADSDALFGSLDKDIEFDHKPEPLFEQIVAGTAKHVGOELPMKPVLPV 289
OY 64 EQLQARLNAAGLNAQPSO--ALDVVNFDDQSPRIGSOPRLDMSVTEETPLSLSE 122
Db 290 AEEPAQVPAEELQTPDDPDLAVT-----ESP-----EVELVDEE 327
OY 123 LFAQESTKGINPNDYIPREYQGEQPNSEVVVPTLEPEKPLIKRLYLRFNDG--VNKV 180
Db 328 ATEESTEVG--PTE-----EGPTELDATP-----EDGFRIRRN 360
OY 181 PRKAKFYSSGGGFSATSSHOKT-EPTANIKAALEDITQESAMDLNLSIPRLQOTAL 239
Db 361 CRRNRNNEGEEETEVAEEGSEETPEGEELQAPED--DFALD-GTTLFEETETAE 416
OY 240 VAARAVGYDIDESTIRNSIGEVDVLIHDGE-----PVIYDRAVVEVREGADDKAF 292

Db 417 GEETVEGEETVE-----GEETVEGEAAEGPEELQATPEDDFQLEEPSGESE----- 463
OY 293 TTVADVEPPLIGVFHNGKYEYTKNLIENASAEHGYDGRWLDKSDVLTLPD----- 345
Db 464 -----GEGEGEGEGEGEALVAVPVAAE-----PVEVVTPAQPVKPMVA 501
OY 346 -TADVSLIYD-TGTQYRFDEVFPTIDPKTNQLTTPDP-----KLPAKRELLQELLVNM 398
Db 502 PTADETLFVQILNDLTYADITSF--EPLRKQILKDPDAGEAVTVPSKAPVOVPVAVGP 559
OY 399 GEAVNIQAVRALNNDLIATRYFMKAVNT-----EIVF-----PREQIQDVS 441
Db 560 AQEVPTEELMQLED-----DFLEGTAAPREGSELVLEGGEPTEREPREGETTEVEVP 614
OY 442 FEOSSSRTEPQAVDESTLEPVIEYV-----ELDTGLIMDS-----PIEFSASNLI 488
Db 615 EEELEATPEDDFELEETGEVEEVEGETABGEVEEVPAAVEEVEEVPAAVEEVEEVP 674
OY 489 QDKLNTVAARARLYDPDQDRLVLAINHDDGVNRSILGRISDAVSAVARAI--LPDESEN- 545
Db 675 PEEVEEVPAAVEEVEEVEEPEE-----VEEVEEVEEVEEVEEVEEVEEVEEVEEVP 719
OY 546 ---EVIDLP-----ERTALANRKTPAVDYQSKR-----VPLVFEVASDKPR 583
Db 720 EEEVEEVPAAVEEVPAAVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVP 779
OY 584 DQOIGIG 590
Db 780 KDELGSG 786

RESULT 37
US-09-092-458-2
Sequence 2, Application US/09092458
Patent No. 6231861
GENERAL INFORMATION:
APPLICANT: Barnwell, John
TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby and Darby
STREET: 805 Third Ave.
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-7513
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/092,458
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/719,821
FILING DATE: 09/30/96
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 5986/17686053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)527-7700
TELEFAX: (212)753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1018 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

```

:
:   HYPOTHETICAL: YES
:
:   ANTI-SENSE: NO
:
:   FRAGMENT TYPE: C-terminal
:
:   ORIGINAL SOURCE:
:
:   ORGANISM: Plasmodium vivax
:
:   IMMEDIATE SOURCE:
:
:   CLONE: pPMB3.3.1
:
US-09-092-458-2

Query Match
Best Local Similarity 18.6%; Pred. No. 0.16;
Matches 124; Conservative 100; Mismatches 227; Indels 216; Gaps 31;

QY 30 AQON-NPAN-----IINHPADRAINQAKGN-----PPVLLTP 63
DB 230 AQENVOPADSNDAFGSILDKDILFDHIKDFEPLFEQIVAGTAKVYTGOLPKMPVPLPV 289
QY 64 EQIQAURLNAGLAKPQSQ-ALDVVNDDQSPISRIGOSPPLGLDMSVIEETTPLSLEE 122
DB 290 AEEPAOVPAEELDATPDDFDALVT-----ESP-----EEVELVLDDE 327
QY 123 LFAOESTEMGINPDYIPEYGEQPNSEVVVPTLEPEKFGLIKRLVYARLFNQG--VNKV 180
DB 328 APEEESTEVG--PTE-----EGPTLELATP-----EDGFRIIRN 360
QY 181 PRKAKFYOSSOGSETSAIGSSHOQT-EPYANIKALEDITOSAMDLSIPRLROTAL 239
DB 361 CRRNRNRNVEGETEEBAEAGEVESEETPEGELEATPED--DEALD-GTTLEETETAE 416
QY 240 VVARANGYDIDLSIRNSIGEVDYIHDGE-----FYIDRYAVEYRGADDKAF 292
DB 417 GEETVEGEETVE-----GAETVEGEAAEGEELEATPEDDFOLEPSESGE----- 463
QY 293 TTVADEVPILGIDVFHNGKETKKNLIENASAEHGYFDGRWLDRSVDVLLPDN----- 345
DB 464 -----GEGEGEGEGEGEALVAVPYAA-----PVEVTPAOPVKPMVA 501
QY 346 -TADVSLIYD-TGTQYRFDEVEVFTIDPKTNQTLTDPD-----KLPRKRELLBQLLVNM 398
DB 502 PRADETLFVIDLDNDLTADITSF--EPLFKQILKDPDAGEAVTVPSKEAPVQVAVVCP 559
QY 399 GEAYNQAVRALSNDLIARIFYMVMVT-----EYVF-----PEREIODQVS 441
DB 560 AQEVPTLEEMLOQED-----DELEGTAAPEBEGELVLEBEGPTPEBREGPTPEEVP 614
QY 442 FPOSSSSRTEPAQVDESTLEPIYETV-----ELTDGILMDS-----PIEFASNLI 488
DB 615 EBELEATPEDELEPTEGEEVEYVGEETAGEVEVEYPAVEVEVEYPAVEVEVEY 674
QY 489 ODKLNLVAAKARHLXDMDDRYALAINHDDGVNRSILGRISDAVASAVARAI--LPDESEN- 545
DB 675 PREVEVEVPAVEVEVEVEVEE-----VEEYVEVEVEVEVEVEVEVEVEVEVEVEV 719
QY 546 ---EVIDLP-----ERTALANRKTADVYQSK-----VPLYVFVASDKPR 563
DB 720 EEVEEEVPAVVEVEVEVPAVEVEVEVEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 779
QY 584 DGOIGLG 590
DB 780 KDELGSG 786

RESULT 38
US-08-769-309A-5
: Sequence 5, Application US/08769309A
: Patent No. 5741890
:
: GENERAL INFORMATION:
:
:   APPLICANT: Scott, John D.,
:   APPLICANT: Nauer, Brian J.,
:   APPLICANT: Klauack, Theresa M.
:
:   TITLE OF INVENTION: Protein Binding Domains of Gravin
:
:   NUMBER OF SEQUENCES: 24
:
:   CORRESPONDENCE ADDRESS:
```

```

:
:   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
:
:   STREET: 6300 Sears Tower/233 South Wacker Drive
:
:   CITY: Chicago
:
:   STATE: Illinois
:
:   COUNTRY: United States of America
:
:   ZIP: 60606-6402
:
:   COMPUTER READABLE FORM:
:
:   MEDIUM TYPE: Floppy disk
:
:   OPERATING SYSTEM: IBM PC compatible
:
:   SOFTWARE: Patentin Release #1.0, Version #1.30
:
:   CURRENT APPLICATION DATA:
:
:   APPLICATION NUMBER: US/08/769,309A
:
:   FILING DATE:
:
:   CLASSIFICATION: 435
:
:   ATTORNEY/AGENT INFORMATION:
:
:   NAME: No. 5741890and, Greta E.
:
:   REGISTRATION NUMBER: 35,302
:
:   REFERENCE/DOCKET NUMBER: 27866/33451
:
:   TELECOMMUNICATION INFORMATION:
:
:   TELEPHONE: 312-474-6300
:
:   TELEFAX: 312-474-0448
:
:   TELEX: 25-3856
:
:   INFORMATION FOR SEO ID NO: 5:
:
:   SEQUENCE CHARACTERISTICS:
:
:   LENGTH: 1780 amino acids
:
:   TYPE: amino acid
:
:   TOPOLOGY: linear
:
:   MOLECULE TYPE: protein
:
US-08-769-309A-5

Query Match
Best Local Similarity 19.2%; Pred. No. 0.42; Length 1780;
Matches 171; Conservative 129; Mismatches 294; Indels 298; Gaps 44;
```

```

QY 23 LMTSQALAQ-----ONNPANIIINHPAH-----DTAINQAKAGNPVLLTEQIQAURLNAA 73
DB 930 LTFEEVLEREVIAEEBPPVTEPLPENREARGDIVVSEAE-----LTFEAVYAAETAG 982
QY 74 GLNAKPQSQALDVVNFPDQ--SPISRIGOSPPLGLDMSVIEETTP----- 118
DB 983 PLSSEGTESAABETTEMVSAVSQTLTD--SP-----DTTEATPVOVEGEGVPDIEQE 1035
QY 119 -----SLEELFAOESTMGIT-NPNDYIPEYO--GEQPNSEVVVPTLEPEKPKLIK 166
DB 1036 RRTQEVLOAAVEYKESQPLGTGEPEDVLOPVORAEAPEREOA-----EASGLK 1087
QY 167 RLVARLFNDGVNRPRLKAK---FYOSSOGSET-----AIGSSHOQT-----E 207
DB 1088 E-----TDVVLKVDAGQEAETEFPTQKVVGGTTPESEFKAPOVTESESSELVTTQOAE 1141
QY 208 PYANIKAALEDITOSAM-----DLNCSIPRLROTALVAVANGYDIDLSIR 256
DB 1142 TLGVKS--QEWMEQOALIPDSVETPTDSETDGSP-----VADFPAPGTOK 1187
QY 257 NSICEVDYIHDGEVYIYIDYRAVEYRGEGADKAPTTVADEVPLIGVFHNGKETK 316
DB 1188 DELIVE--IHEENE-----VHLVPVRGTEAE--AVPAOKERPPAPSPVFQOE--ETKE 1233
QY 317 NLLENASAEHGYFDGRWLDRSVDVILPDMTADVSLIYDT-GTQ-----YRDEVEVFF- 367
DB 1234 QSKMEDTLEH-----TDKEYSV-----EYVSIISKTEGTQDAQYADDKTKDVPFEE 1280
QY 368 ---TDPKTNQTLTDPDKL-----PVKRELE 391
DB 1281 GLEGSID--TGITVSRREKYTEVALKEGTEAECKKDALQLSHAKSPSPSVEREMVY 1337
QY 392 QL-----LTVNNGEAYNLOAVRALSNDLIATRYFMVMVTEIYFPERE----- 433
DB 1338 OVEREKTEAPPHVNEKEKLEHETAVTV--SEEVSKOLLQTVNPIIDGAKEVSSLEGSPP 1395
QY 434 -----QIQNDVSEFOSSSSRTEPAQVVESTLEPIETVELDGIILMDSPI 480
```



```

: COUNTRY: USA
: ZIP: 92121
: COMPUTER READABLE FORM:
: MEDICAL TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/436,664
: FILING DATE: 08-MAY-1995
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/394,232
: FILING DATE:
: APPLICATION NUMBER: 08/307,410
: FILING DATE: 16-SEP-1994
: APPLICATION NUMBER: 08/222,612
: FILING DATE: 16-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Fisher, Carlos A
: REGISTRATION NUMBER: 36,510
: REFERENCE/DOCKET NUMBER: GP94003.CP2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-546-2807
: TELEFAX: 619-546-7929
: TELEX:
: INFORMATION FOR SEQ ID NO: 34:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 876 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: Internal
: ORIGINAL SOURCE:
: US-08-436-664-34

```

```

Query Match      2.6%; Score 123.5; DB 2; Length 876;
Best Local Similarity 19.2%; Pred. No. 0.14;
Matches 193; Conservative 147; Mismatches 358; Indels 307; Gaps 51;

```

```

QY 8 ANSEMPVAL-----AAY-LPLMTSOLAQONNPANIINHPA-----HDTAIN 50
DB 14 AYRAFPALPLHNDKGIHTNAVYGFTMLNKLIAEE-QPTHILVAFDAGKTTFRHER-FQ 71
QY 51 QAKAG---NPPVL-----LTPEIOARLNAAGLNAPQOSALDVNFDOSPISRIGESQ 102
DB 72 DAKGGHQGTPELSEQFPPLRELLKAV-----RIPAYELDHYEADDIIGTMARA 121
QY 103 PVLGLMSVI---EETTPLSLEELFAQESTEMGINP-NDYIPEQSGQPNSEVVPPTLE 158
DB 122 ERGEGFAVKYISGRDQLTOLASPOV-TVEITKGGITDIESTTPE-----TVEKYGLT 172
QY 159 PEKPGILKRLYALFNDGVNKPRLAKFYQSSQSGETSAIGSSHOKTEPYANIKALED 218
DB 173 PEQIVLKLGMGDK-SDNIPGVPGI-----GEKTAV-----KLLKQFIVENVLAS 217
QY 219 ITQESAMDLSIPLRLQRLAVAR-AVGYYIDLSIIRNSI---GE---VDYIHDLG 270
DB 218 IDEIKGEKLEKRLROYRDIALLSKQLAALCRDAPVELITLDIYKGGEDREKVVALLFQELG 277
QY 271 EPIYIDYRAVEV-RGE---GADDKAFTTYADE-----VPLLIGDVFHHGKYETKKN 317
DB 278 FOSFLDKMAVOTDEGEKPLAGMDPAIADSVTDLADKALVAIVEVGDNYHHAP----- 331
QY 318 LIENASA-EHGYFDGRWLDRSVVILPDNT-ADVSLIYDTGTQYREDEVVFTI-DPKTN 374
DB 332 IYGIALANERGRF-----FLRPETALADPKFLAWLG-----DETKKKTMFDSKRA 376
QY 375 QLTTPDCKLPVKRELLLEOLLTVNMGEAYNLQAVRAISNDLIATRYENMVNTEIVPEREQ 434

```

```

DB 377 AVALKMKGIELRGVVELLLA-----AYLDDPAQA-AGDVAAVAKMH-----QYEA 421
QY 435 IONDOVEEQSSSSRTEPAQVDESLPEVLETFVETLDTGILMDISPIESASNLIQDKLNL 494
DB 422 VRSDAVYVK-CAKRYTP--DEPTL----- 443
QY 495 VAAKARHLVDPDDRVLAINHDDGVNRSILGRISDAVASAVARAIIIPESENE-----VI 548
DB 444 ---AEH-----LARKAAAIWALEEPIMLDELRRNEQRLLTTEL 477
QY 549 DLPERTALANKRTPADVYQSKVPLVYFVASDKRRDQOIGIGMSDPTGRLTYTFEHLI 608
DB 478 EQLAGILANNE-----FTGVKVDYTKRLQDM-----GAELTEQLQ--AV 514
QY 609 NRDOYO-AGAEPLRSEDDKGVKLYATKPLSHPLNDOLRATLGYQOEYFGHSTNGFDLSTR 667
DB 515 ERRIYELAGQEFNINSRKQ-----LGVTLDKL-----QLPYLKKTKTGYSTISAD 559
QY 668 TLE-----HEISRSII---QNGGWNRTY-----SLRYRLDKLTKQAPPET----- 704
DB 560 VLEKLAPHHETVEIHLHYRQLGKIQSTYIEGLKLVHPVYGVKVTHTFNOALTYGRLLSSV 619
QY 705 ---WQDLPVDFVNGKPSQDEALLAG-----FTGVKVDYTKRLQDM-----VAVHKTVAONLVNPMKGYR 744
DB 620 EPNLQNIPIRLLEGRKIRQAFVPSPPMLIFPADYSOIELRLVLAHIAEDDNLIEAFRGL 679
QY 745 QRYSELEVSSGLVSD---ANMAIARAGISGVSEFGDNAGVSNRAHQMTGSIQAGYIMSD 800
DB 680 DIHTKTMDFHSEEDVTANMRQAKAVNFGIYGISDGLAONLITTKREAAEPI--- 736
QY 801 NFNHVPRRLRFFAGDOSIRGYAHSLSPISDKGYLTGGVLAAGTAEYNYEFKKDLRLA 860
DB 737 -----ERYFA-SFPGVQYQMDNIYQEAQKQGYV-----TLLHRRRYLPDITSR 779
QY 861 VEGDIGANVYDKGFTNDKTKIGAGVGVNMAVSPYGVYRVNVAIVGKKE 905
DB 780 NFN-----VRSEFAERTAMNTPIOGSAADIKKAMIDLVSRLREE 818

```

```

Search completed: April 28, 2003, 16:24:54
Job time : 65 secs

```



```

QY 50 NQAKAGNPVLITPEQIQARLNMAAGLNAK--POSALDVYAFNDDOSPISRGEQSPPLGLD 100
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 24 NAASKDDPSIVVLEPFOINGSSNDEELQTELEPMLATALKN-----KG 65
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 109 MSVTEETPLSLEELFAQESTEMKINNDYIPEYQSQPMSVYVPTLEPEKGLIKRL 160
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 66 FRIYIPKSAIINTL--LYKQNTSOLINISPAKRV---AQLHADYVV----- 104
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 169 YARLFNDGVKAVPRIRAKKFFYQSSQSGTSAIGSS---HQTEPPIANKALLEDITQESAM 225
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 105 YGSEFNQTGEN-----FSIDSLRLDPTGVASRPVIEKPKFNNELINAVTELAERIS 155
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 226 DLWGSIPLRQTAIVAAARANGYVDIDSLTRNSGGEVDYIIF-----DLGEPVY 274
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 156 --NGLIKK--NTIDAVRIHGLEKVLDPDVIITRLINKGDHFDHAKINAEIKKIWEIG--VF 210
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```

333 275 IDYRA-VEVREBGADKFAFTTVADEVLLIGDVFHCKYETKKNLLENASAEHCYDGRW
 Db 211 SDVASAIESEEGE---RLVFTVOEKFRITDVVQSKAVSIDILAMSSKKG-----261
 QY 334 LDRSV--DVLLPDMTADVSLIYDGTGYRDEVEVFFPIDKTNQTLTTPDKPLVYKRELL 391
 Db 262 ---SVISDRLLSQIOIKITTDYRKEGYVLAE--VNYEIKREKEMTSSA-----T 304
 QY 392 QLLTYNMGAEVNLQAVRLSLMDLIATRYFNKNVTEIYFPER-----EDION 437
 Db 305 LLTYVNECKKUYIKIDVREGLLETKAN--TLKKELALTERNELFSLWTTGTGVLRREYLER 361
 QY 438 DOVSEEOSSSRTEPAQYDESTLEP-----VIEVTELDGILMDISPESASNLIQ 489
 Db 362 DSIAI---SAYAMNHGYVDIQVAPSEVTFNFKNGVITFVRKCEGRYRIKIGIDKGDIIET 418
 QY 490 DKLNLVAAKAARHLTYDMPDRYLAINHDDGYNRSILGRIIDAVASAVARAILPDESENE--- 546
 Db 419 NEOLLKVTYKIDQHNKNYEOYFSLSMQWD--VKALTDFYSDGYAFAEVLD--ETTKEDEDA 475
 QY 547 -----VIDLPERPAL-----ANRKTPEADY-----566
 Db 476 TIDYTFLLDKKQKQKFLRIIYEGCTRTRDVYILRELRLADGDLFNGOHLRSECLNRLG 535
 QY 567 -----OSKRVPLVYFVASDKPRDQOIGLQMGSDGTGRIVTK--FEHNLINR 610
 Db 536 YFNQVDYDTLPTGKDEVEDYLLVYK--QEARGTATGGVGVSTHSKFGVSSISEBNLMGK 593
 QY 611 DGYGAGANLRLSEDKKCYKLATATPRLSHPLINDQIRATLQGYEYFEGHSTNGFDL----- 664
 Db 594 -GYLSTIEGFITSSKSSSIDLSEFTPRVYTD-----FGFSNNIYTLRDEWD 639
 QY 665 -----STRLEHEISKSIIQNGMNRYSLSRYRLDKLTQAAPPETWQDLPVDFVNGKP 717
 Db 640 FRKTYGDTILRHPRI-----GEYSSTIFGVGRIDQRYRLYDIPSRAPSYLDY-QGKN 690
 QY 718 SQEALLAGVAHVAKTVADNLVPMHGYQORISLEYGSSGLVSDANMAIARAGISGVYSFGD 777
 Db 691 TSSVSGGFETFDST--DSREPRSGHIAKLIVEGGGGGLGNGNFFKPIAELOGFYSI-- 746
 QY 778 NAVSNNRAHQMTGCIQGYIMSDNFNHPYRLRPFAGCGDSIRGVANDSLSPISDK-GYL 836
 Db 747 ---SRSKNHIIHMTTRAGAAAKNKKRPVPYDPRFFIGTIDISIRGYDIEDLAPADPRFGDE 803
 QY 837 TGGQVLAVGTAEVYEFPAKKDRLAV--FGDIGNAYDKGFTND---TKIGAGGVWRWASP 890
 Db 804 IGGGRMAFLNLEYIMTQPELIGLALVPFYDIFGFTDSVQTSNPFSSKLQSGYLELRWSP 863
 QY 891 VGVOR 895
 Db 864 MGDLR 868
 RESULT 2
 US-10-080-505-13
 : Sequence 13, Application US/10080505
 : Publication No. US20030073166A1
 : GENERAL INFORMATION:
 : APPLICANT: St. Gene, Joseph W.
 : TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
 : FILE REFERENCE: A-59941-1/RET/DCF/DHR
 : CURRENT APPLICATION NUMBER: US/10/080,505
 : CURRENT FILING DATE: 2002-02-22
 : PRIOR APPLICATION NUMBER: US 08/296,791
 : PRIOR FILING DATE: 1994-10-25
 : PRIOR APPLICATION NUMBER: US 09/839,996
 : PRIOR FILING DATE: 2001-04-20
 : NUMBER OF SEQ ID NOS: 58
 : SOFTWARE: PatentIn version 3.1
 : SEQ ID NO 13
 : LENGTH: 1436
 : TYPE: PRT
 : ORGANISM: Haemophilus influenzae

[illegible]

```
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR APPLICATION NUMBER: 2001-03-21
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5635
LENGTH: 2368
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5635
```

```
Query Match      2.9%  Score 137.5; DB 10; Length 2368;
Best Local Similarity 17.4%  Pred. No. 0.89;
Matches 202; Conservative 146; Mismatches 423; Indels 387; Gaps 47;
```

```
QY 2 SKPVLEFANSFPMVALAAYPLMTSQLAQONNPANT-----INHPAHMTALNOAKAGN 56
DB 471 TREVLFPDPS-----LKLSTKYVNAVANIDTPKKNIDENEKLYRTASDIVINMO--- 518
QY 57 PVLVLT--PEQIOARLNAAGLNKKPSQA-----L 84
DB 519 PEVTLTADDFSVAVENKMKDALQOVNSQVNSHYTTASTAEVKKLQOQADNIILEDANHY 578
QY 85 DVVNFDDQSPI-----SRIGBOSPPGLDMSVIEETTPULSELEFAQES-----TE 130
DB 579 ETANRASQAIDGLVTKQIAALIDNQAAIAELDAKAEKVTAQOSKKVTQDEVALVTK 638
QY 131 MGINPNDYIPRY-----OG---EOPN-----SEVVVPPTLEPE-KPGLIKRLIYAR---- 171
DB 639 INNDKNNAIAELINKOTTAQGVTEKKONGIAVLDDQVITPTVKQAQODITQAVTTKKOOI 698
QY 172 -----LFDNGVKNVPRLKAKEYOSSQSG-----ETSAIGSSHOKTEPYANIK 213
DB 639 KKSNASLDEKDVANDKICKIEFKAIKDIDAFITNAQOVAIKTKALINDINO-TAPATTK 757
QY 214 -AALF-----DITQESAMDINGSTIPRLQGTALVAARAVGYDI--DLS 253
DB 758 AALLEEFDEVQAOIDQAPLNPDTTNEEVAE--AIERINAKVSGVKAIEMTTTQODJE 814
QY 254 IIRN-----SIGEDVVIHIDLG 270
DB 815 RVKNEIEIKIENTITDSTQTKMDAYEVKQOATARTKQNTATVSNATNEEVAEADAVAAQ 874
QY 271 EPPYIDYRAVEVREGADDKAFTTVADEVPLT-----IGDVFHGGKYEKKNLJENA 322
DB 875 KQGLHDIOVNVSKQEVADTK--SKYLDKIMAIQOAKKPAADTEVENAYNTRKQEIQNS 932
QY 323 SAE-----HGFDGMLDRBSVDVILPDDNTADVSLIYDTG-----TOYRFD-- 362
DB 933 NASYTEEKQAAVTELDTKQOEARFTNLDAANTNSDVTAKDNGIAIINOVAATTKKSDAK 992
```

```
QY 363 -EYVFTIDPKT-----NQLTDDPKLPYKRELEBOLLTVNKGAVYNQOAVRALSNDLIA 416
DB 993 AEIAQKASERKTAIEAMNDSTEEQO--AAKDQVQAVVTANADIDN-----AAANTVDVN 1046
QY 417 TRYFNMTVELIVPE-----REQIONDOVSFFOSSSSRIE-----PAQVEDS 458
DB 1047 AKTTNEATIAITTPDANVKPPAKAIDAKVOAQETAIIDANNGATTEEKAARKQOQVTEKT 1106
QY 459 TLEPVIE-----TVELTDGILNDISPIEF-----SASNLIDQKLNVAKARHLYDM 505
DB 1107 TADTAIDGAHTNAEVEAKAKNEIAKIEAIOPATTTTKDKAKAIAIKANEKRTAIAQOTDI 1166
QY 506 PDDRYLAINHDDGVNRSITLGRISDAVSAVARAILPDESENEV-----IDLEPRTALA 557
DB 1167 TAEETIAAN-----ANVDNAVTVQANNNIEAANSQNDVDAKTTGCEASIDQVPTV 1216
QY 558 NRKTPADYQSKKPYLVFVVASDKPRDQOIGLGMGSDGTGLVTKFEHNLINRQGYQA-- 615
DB 1217 NKKATAVTDAKNNI--TAAITDQGVPTAKDAGKNSIOSTOPATAVYSMAKN-DVDQAVT 1272
QY 616 -----GAEPLRLEDKKGVKLYATKPLSHPLNDQLRATLGVOQEVFGHSTNGPDL 665
DB 1273 TQNGAIDNTGA--TTEEKMAADVLVAKAEKAYQDILNA-----QTTND----- 1315
QY 666 TRTLEHISRSIIONGWNRTYSLRYRLDKLKTQAPPE-----TWODLPVDFVNGKP 717
DB 1316 -----VTQIKDQAVADVQIGTADTTITKDAKDELATKA 1348
QY 718 -SQEALLAGVAVHKTVDADNLVNPMBGRQKRSLEVGSSGLVSDANMAIARAGISGVYSPG 776
DB 1349 REQKALTAQTADATTEEKEQANOQ-----VDALETQGMONINEMAOSID 1391
QY 777 DNAYGSNRHQMTGGIOAGYIWSDFNHPVRLRFAGGDSIGVAYHDSLPSIDMGL 836
DB 1392 DVMTAKMNAIQALDIPLOASTVDKTN-----ARAEMLTEMONKKTIELLINNETTNEEG-- 1444
QY 837 TGGQVILAVGTAEXNYEFMKDLRLAVFGDIGNAYDKGFTNDTKIGAGVGVWASPVQGV 896
DB 1445 -----NDI-----GPVRAAYEEGLNN-----INAAITTGQVTT 1472
QY 897 DVATGVKE-----EGNPVK 910
DB 1473 AKDTAVQKVOQLHANPYK 1490
```

RESULT 4
US-09-815-242-12389
Sequence 12389, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 12389
 LENGTH: 2368
 TYPE: PR
 ORGANISM: Staphylococcus aureus
 US-09-815-242-12389

Query Match 2.9%; Score 137.5; DB 10; Length 2368;
 Best Local Similarity 17.4%; Pred. No. 0.89; Indels 387; Gaps 47;
 Matches 202; Conservative 146; Mismatches 423;

2 SKPLFANRSFMPALAAVLPMTSQLAQQNNPANI-----INHPADHAIINQAGN 56
 Db 471 TKEVLFPPKS-----LALSYKVNANIDTPKIDPFNEKLTFRASDIYINNAQ--- 518
 QY 57 PPVLT--PEQIOARLNAGLNAPSOA-----L 84
 Db 519 PEVLTADPFSAVEMNKDALQQVNSQVNSHYTTASIAEYKCLKQAADNLINEDANHY 578
 QY 85 DVNFDOSPI-----SRIGOSPPLGLDMSVIEETTPLSLEELFAOES-----TE 130
 Db 579 ETANRASQAIDGLVTKQAALIDNQAIAELDAKQEVTAQAQSKVYTDQVVALATK 638
 QY 131 MGINPNDYIPEY-----QG--EQPN-----SEVVPPLEPE--KPLIKRLVAR--- 171
 Db 639 INDKNNIAIEINQOTQAQVTEKDNGLAVLDQVITPVPKAKQOIIQAVTRKQOI 698
 QY 172 -----LFDGYNKVPRLKAKFYQSSQG-----FTSAGSSHQTEERYANIK 213
 Db 699 KKSNAIODEKDVANDKIGKIEKAIKIDATTAQVNAOYEAIRKTAINDINO--TAATATK 757
 QY 214 -AALE-----DIQESAMDLNGSIPRLQOTALVAARAVGYDI--DLS 253
 Db 758 AAALPEPEVVOAQIDQAPLNDPTNEEVAE--AIERIMAKVSGVAIAEITTAQDLE 814
 QY 254 IIRN-----SIEVDVITIHDLG 270
 Db 815 RYKNEISKIENITDSTQTKMDAYNEVQOATARKTONATVSNATNEEVAEADAVEAAQ 874
 QY 271 EPVYIDYAVEVREGGADKATFYADEVPLL-----IGDVHHGKYEKKMLIENA 932
 Db 875 KQGLHDIOVAKSKQEVADTK--SKVLDKINAIQOAKKPAADTEVENAYNTRKQEIQNS 932
 QY 323 SAE-----HGYPDGMILDRSVIILPDNTADVSLIYDTG-----TOYRFD-- 362
 Db 933 NASTTEEQAAVTELDTKQKQARTNLDAANTNSDVTAKDNIAALINVOQAATTKKSDAK 992
 QY 363 -EVEFTDPTK-----NQLTTDPKLPVKRLEHQLLTVNKGAVNIQAVALSNDLIA 416
 Db 993 AEIAQKASERKTAIEAMNDSTTEEQO--AAKDKVQAVVATANADIDN--AAATVDVN 1046
 QY 417 TRYFMVMTVEIYFPE-----REGIONDOVSFEQSSSSTE-----PAQVDS 458
 Db 1047 AKTTNEATIAATTPPANAKPTAKQALADVQAQETAIIDNNKATTEEKAQAQVOTEXT 1106
 QY 459 TLEPVIE-----TVELTDGLMDISPIEF-----SASNLIDQKLNVAAKARHLXDM 505
 Db 1107 TADTADGAHNAEVEAAKNMAIEAIOPATTTKDNKAKQALANKRANKRAIAIQTOI 1166
 QY 506 PDDRVLAINHDDGVNRSLIGRISDAVSAVARAILPDESENEY-----IDLPERTALA 557
 Db 1167 TAEETIAAN-----ANDNAVQTQANNNIEANSQNDQADQAKTTGEASIDQVTPV 1216
 QY 558 NRKTPADVYOSKKVPLVYFVASDKPRDQIGLGWGSQDTRLVTFEHLILNRDGOA-- 615
 Db 1217 NKKATAVTDAKNNI--TAATDNGVDTRAKDAGKNSIGSTQPATYAVKSNAKN-DVDAQAVT 1272

QY 616 -----GAELRLSHDKKGVKLYATKPLSHPLNDQLRATLGGYQDEFGHSTNGFDLS 665
 Db 1273 TONQALDWTGA-----TTEEKNAAKDLVLKAKEKAVQDILNA-----QTND----- 1315
 QY 666 TRTLEHETSRSTIONGKNNRTYSLRYLKDCLKTOAPE-----TWQDLPVDPVNGKP 717
 Db 1316 -----VTOIKDQAVADVQGITADTTIKDVAKDELATKA 1348
 QY 718 -SOEALLAGVAVHKTVDNVLNVPMBRGYRQRYSLFEGSSGLVSDANMAIARAGISGVYSG 776
 Db 1349 REQKALINQOTADATTEEQANO-----VDALITGQNGQNIENAGSID 1391
 QY 777 DNAGSNRAHQWGTGIGQGYIWSDNFNHVPYRLREFPAGDOSIRGYAHDSLSPIDKGL 836
 Db 1392 DVNTAKDNAIDAIDPLQASTDYKTN-----ARAELLTEMQNKITELINNETTNEEG-- 1444
 QY 837 TGGQVLAVGTAEYNEFPKDLRLAVFGDIGAIDKGFINDTKIGAGVGRNASPVGQVRV 896
 Db 1445 -----NDI-----GPVRAAYEEGLNN-----INAAFTTGQDVTT 1472
 QY 897 DVATGVKE-----EGNPIK 910
 Db 1473 AKDTAVQKQQLHANPVK 1490

RESULT 5 US-09-815-242-11571

Sequence 11571, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA 011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 11571
 LENGTH: 2890
 TYPE: PR
 ORGANISM: Helicobacter pylori
 US-09-815-242-11571

Query Match 2.9%; Score 136; DB 10; Length 2890;
 Best Local Similarity 19.4%; Pred. No. 1.5; Indels 318; Gaps 52;
 Matches 198; Conservative 145; Mismatches 361;
 QY 55 GNPVLTPEQIOARLNAGLNAPSOQALDVNFPDOSPIRIGOSPPLGLDMSVIEE 114
 Db 1675 GAPEITVRNEK-----RMLQEAADVLFQDGRSTNAVKGANKRPL----- 1713

Db 1765 -----RLAAKHLEQALNOAIDQINHA-----DRTAOVNQDSINAQINISIKRPAT 1810
QY 532 SAVARAI-----LPDESENEVIDLPE-----RTALNRKTPADYV 566
Db 1811 TVKATALQOIONATNKINLIKANNEATDEQONIAIAQVEKELIKAKOQIASATYNAV- 1869
QY 567 QSKRVPLVYFVASKPRDQIGLWGSPTGTRLYTKFEHNILNRDGYQAGALRLSEDK 626
Db 1870 -----AYLHDEKNE-----IREIE-PVINRKASAREQLTTLFNDRK 1905
QY 627 GVKLYATKPLSHPLNDOLRALTYGOEYFGHSTNGFDLSTRTLEHISRTIIONGWNRT 686
Db 1906 -----QALEANIQATVEERNSTLAOLQNIYDAIGQIDORS-----NAQVKT 1949
QY 687 YSLRYRLDKLKTQAPPEWODLPVDFVNGKPSQEALLAGVANHKTVDNLVNP---MRGY 743
Db 1950 ASLNL-----QTIHDLVHPIK-KPDAE-----KTINDLARVTALVQNY 1988
QY 744 RQRYSLLEVSSGLVSDANMAIARAGISGV 772
Db 1989 RK-----VSNRNKADALKAITATL 2006

RESULT 7

US-09-815-242-12913
; Sequence 12913 Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlseen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 12913
; LENGTH: 2186
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-12913

Query Match

Best Local Similarity 19.0%; Score 135; DB 10; Length 2186;
Matches 154; Conservative 110; Mismatches 301; Indels 244; Gaps 35;

QY 24 MTSQALAOQNPNPANIINHVAHDAITAIQAKAGNPVLLTFEQIARLMAAGLN----- 76
Db 1382 VASQALAKEKEKRL-----AATDQA-----QINSQVNOAATNCVSIAIKI 1420
QY 77 ---AKPQSOALDVNFDQSPISRIQEQSPPLGLDMSVIEETTPLSLE---ELFAQEST 129

Db 1421 IQPETHVPAAREKTIN-----QKANEIRAKINODKEATAEEROVALDKINEFVNOAQT 1473
QY 130 EMGIN-PNDYIPREYQGEOPNSEVVPPTLEPEKPGILKRLYARLFENDGVNKPRLKAFY 188
Db 1474 DITNRTNOQVDDTISQALDSTALY-----TPPHIVAAAR---DAVKQ--QYEAARR 1521
QY 189 OSSGGETSAGSSHQKTEPEPVANITKALEDITQESNMDLNGSIPRLROTALVAARAVGY 248
Db 1522 EIEQAEHATDEEKQVALNOLANNEKRALQNIQAI-----NNQVKRETNGIATLKGV--- 1575
QY 249 DIDLSTINNSIGEVDVYIHDGEPEYIDYRAVEBGEADKAFITY--ADVEPLLIGV 306
Db 1576 -----QPHIVIKP--EAQQAIKASENQVESIKDIPPHATVDELDAQNLISPT 1621
QY 307 FHHGKETKKNLIENASAEHGYFDR-WLDRSDVYLPD---NTADVSLYDPTGYRED 362
Db 1622 LKQAQEE-----IENTNQDAATDVNRQTIKAIEDQIKPRVRKRAALDSIEENNNK--- 1672
QY 363 EYVFETIDPKTNQLTTDPDKLPEVKRELLFQLLTVNMGEAYNIQAVRALSNDLIATRYFM 422
Db 1673 -----QLDAIRNTLDTTDQERDVAIDTLNKI-----VNTIKNDIAQ-----NK 1710
QY 423 VNTLIVPEPEREQIONDOY-----SPEOSSSRTEP--AQVESTLEPYLETVELTD 471
Db 1711 TNAEYDRTETDGNINIKVILPVQYKPARQSVGYKAEQNALIIOQSLSTEE----- 1764
QY 472 GILMDISPIEFSASNLIDKINLVAKARHLYDMDDRYLAINHDDGVNRSILGRISDAV 531
Db 1765 -----RLAAKHLEQALNOAIDQINHA-----DRTAOVNQDSINAQINISIKRPAT 1810
QY 532 SAVARAI-----LPDESENEVIDLPE-----RTALNRKTPADYV 566
Db 1811 TVKATALQOIONATNKINLIKANNEATDEQONIAIAQVEKELIKAKOQIASATYNAV- 1869
QY 567 QSKRVPLVYFVASKPRDQIGLWGSPTGTRLYTKFEHNILNRDGYQAGALRLSEDK 626
Db 1870 -----AYLHDEKNE-----IREIE-PVINRKASAREQLTTLFNDRK 1905
QY 627 GVKLYATKPLSHPLNDOLRALTYGOEYFGHSTNGFDLSTRTLEHISRTIIONGWNRT 686
Db 1906 -----QALEANIQATVEERNSTLAOLQNIYDAIGQIDORS-----NAQVKT 1949
QY 687 YSLRYRLDKLKTQAPPEWODLPVDFVNGKPSQEALLAGVANHKTVDNLVNP---MRGY 743
Db 1950 ASLNL-----QTIHDLVHPIK-KPDAE-----KTINDLARVTALVQNY 1988
QY 744 RQRYSLLEVSSGLVSDANMAIARAGISGV 772
Db 1989 RK-----VSNRNKADALKAITATL 2006

RESULT 8

US-09-765-272-118
; Sequence 118 Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 1040 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 118:
US-09-765-272-118

Query Match 2.8%; Score 134; DB 10; Length 1040;

Best Local Similarity 19.4%; Pred. No. 0.44; Matches 190; Conservative 124; Mismatches 318; Indels 348; Gaps 53;

QY 24 MTSQALAOQNNANI--INHP-----AHDTAIQAKAGNPVLLTPEQIOARLNA 73
DB 52 LTSQLSAANSQSLSGVGHLEPLKIEGYIGYIKTKKQNDTELSTRT---VDGKYSQ 108
QY 74 GLNAKQSQALDVNFD-----QSPISRIQSGSPPLGLDMSVIEETPLSLELFAOE 127
DB 109 RDSQNSTSTSVSHVADLEMMQGGKVSLOGRSGDDGL-----SEKSIADNLSND 163
QY 128 STEKINPDIPEVQGEQPNSEVVVPPLLEPKRLIRLARLENDGVNKKPKAKF 187
DB 164 SPASQVBOH--PDHGES-----VVRPTV-PEQGNPVA----- 194
QY 188 YOSQSGETSAIGSHQKTEPYANIKALEDI-TQESADLNGSIPRLQOTALVARAVG 246
DB 195 -TTVQSAEEVLAITNDRE-----YKLPLETKGTQEPGHE----- 229
QY 247 YYDIDLSTIRNSIGEVNIIHDLGEPVYIDYRAVEVRG-EGADDKAFTVADEVPLIGD 305
DB 230 -----GEA-AVREDL--PVYT--KPLETKGTQEPGHEGEAAVREPEAYTER 271
QY 306 VF-----HHGKETKNLE-----NASEHGTFDGRMLDRSDVILPNTAD 348
DB 272 LAFKGTQEPGHEGKATVREETLEYEPVATKGTQEPH--EG--ERVVEELP----- 320
QY 349 VSLIDTGTQYFDEVFETIDPTNQLTTPDKPLPVKRELEQLLTVMNGEAYNLQAVR 408
DB 321 -ALEYTRNRRTIQNI-----PYTTEIQDPTLLKNRKRIEQ-----GQA----- 360
QY 409 ALSNDLIATRYFN---MVNTEIVPEREQIQNDQVSFEQSSSRTEPAQVDE---STL 460
DB 361 -----STRTIQYEDYIVGNV-----ETKEVRSREPAVNEVAVKVTL 399
QY 461 EPVITVELTQIIMDISIEFSAS-----NLIQDKLNLVAAKARLLYMPDDRLALNH 515
DB 400 VKKFTVEIT-----NLKVENKKSTIVSYNLIDTTSAYSAKQ-----VFH 442
QY 516 DDGVRSILGISDAVASAARAILPDESENEYIDLPERFALANRKTPADVYQSKVPLV 575
DB 443 GD-----KLVKEVDIENPAKEQVI-----SGLDYIPIYTKTHL 476
QY 576 FVASDKPRDQIGIGMGSDGTGLVY--KFEHNLINRDQYAGABELRSEDKGKYLVA 632
DB 477 -----TYNLGNNENMETETSTQDFQLEYKIE-----IKQIDSELYLG 514
QY 633 TKPLSHPLNDQIRATLGVQGVFGHSTNGFDLSTRLHEIRSRSTIIONGGMNRTYSLRR 692

DB 515 KE-----NDIRYRYSLS-----EAPTDTAK-----YFVKV 541
QY 693 LDKLQTQAPPETWQDPEVFNKGPQSEALLAGVAHKTAVADNLVNMGRYRORYSLEVG 752
DB 542 SDRFEKEMV-----LPVKSTI--ENTDGYTKVAVADQVLEEGL---TDGYKDDYFTVA 589
QY 753 SS-----GLVSDANMAIA--RAGISGYYSFGDN-----AYGSRAHQMTGGIQAGTMS 799
DB 590 KSKAEQPGVYTSFKOLVATAMQSNLSGVYTLASDMDADEVSLGDKQTSYLNGAFGLS 649
QY 800 DNF--NHVPYRLR-----EFAGG---DOSIRGYAHD-----LSPISDKY 835
DB 650 DGKSTAIYDLKPLDPTLNGATVRDIDITVSADSKENVAALAKAANSINNVAVEGK 709
QY 836 LTGGQVL-----AVGTAEVYEFMKDLRLAVFGDIGNAVYDKGFTNDTKIGAGVG--VRW 887
DB 710 ISGAKSVAGLVASATMTVIENSFTGKL-----IANHQDSN-KNDT--GGIVGNITGN 759
QY 888 ASPVGOVRVDVATGVKEEGN 907
DB 760 SSRVKNRVYDALISTNARNN 779

RESULT 9

US-10-160-758-15
Sequence 15, Application US/10160758
Publication No. US20030036076A1

GENERAL INFORMATION:
APPLICANT: EXELIXIS, INC.

TITLE OF INVENTION: CAPS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE

FILE REFERENCE: EX02-089C
CURRENT APPLICATION NUMBER: US/10/160,758

CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US 60/296,076

PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/328,605

PRIOR FILING DATE: 2001-10-10
PRIOR APPLICATION NUMBER: US 60/357,253

PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 16

SOFTWARE: PatentIn version 3.1
SEQ ID NO 15

LENGTH: 4349
TYPE: PRT

ORGANISM: Homo sapiens
US-10-160-758-15

Query Match 2.8%; Score 133.5; DB 9; Length 4349;

Best Local Similarity 17.4%; Pred. No. 4.4; Matches 199; Conservative 162; Mismatches 388; Indels 395; Gaps 50;

QY 52 AKAGNPVLLTPEQ--IQARLNAAGLNKQSQALDVNFDQSPISRIQSGSP--LGID 108
DB 2145 ARDGPPLQSEBEVLTVYNNKSNPLFQSPYKVRVPEPNTTLYTPIILHTQARSPGRLI 2204
QY 109 MSYIEE-----TTPLSLE---ELFAQESTMGINPNDYIPEYQGEOP 147
DB 2205 YNIVEEPEPLMLFTTDKTKGVLYTQGLDYESKTKHFTVATATPAGL-----SFESEA 2256
QY 148 NSEVVV-----PPTLE-----PEKPLIKRLVA-----RLENDGV 177
DB 2257 TVEVLVEDVNDNPPFPQSQALVYTSISEGLPAQTPVLIQLLASDQSGNRDVSQIIVEDGS 2316
QY 178 NKVPRLKAFQY--SSQSGETSAIGSHQKTEPYANIKALED-----IQESAM-----D 226
DB 2317 D-----VSKFFQINGSTGEISTVQELDYEQQHFKVVRAMDGDPPLVGETLVVNVVSD 2371
QY 227 LNSGIRPLQOTALVARAVGYIIDLSIIRNSIGEVNIIHDLGEPVYIDYRAVEVRG-- 285
DB 2372 INDNPEFFQPOVEANVS-----ELATGCHLVKQALIDPDRSDRSRLLEYLLISGNQR 2425
QY 286 -----GADDKAFTVADEVPLLIGV-----FH 308

```
Db 2426 HETINSSGIIIMFNCKKHLDDSSYNLRVGASDGVFRAT---VPVYINTTNANKYSPEEQ 2482
OY 309 HGKYEKKKMLIENASAEHGFEDGRLMDRSVAVILPDNTADVSLIYDTGTQYRDEVEVFT 368
Db 2483 OHLYEAE--LAENAMVGTAKVIDLAIDKDSG---PYGTIDTITINKLASEK-----FS 2530
OY 369 IDEPTNOLTTDPPKLPVKRELLEOLLTVNMGEAVNLQAVRALNSD-----413
Db 2531 INP-NGOIAI-----LQRLDRENSSTE--RVIAIKVMADGGGRVAFCTVILIT 2576
OY 414 -----LATRYFNNVNT-----ELVPPERQIONDOVSFESSSSSRTEPAVD 456
Db 2577 DENNPPQFASETTVTSIQSVNSKDSPIYQVLAVDAGQNAVDY---SVNPDVLKD 2632
OY 457 ESTLEPIETVELTDGIL-MDISPIEFA-----SNLIDKMLVAAKAR-----500
Db 2633 VIEINPYTGAVKVDLSVLENGQFLDEFIKAQDGGPRHMSLVRLQVVPKKYSLPKFS 2692
OY 501 ---HLVMPDPRVLAITHHDDGVNSILGRISDAVSAVARAILPDESENEVIDLPEBTALA 557
Db 2693 EPLTFESAPED--LPEGSEIGIVAKAAQ--DPVYISLVKGTTPESNKNKGVSFLDPDGTGI 2749
OY 558 NRKTPAD-----VYQ-----SKK 570
Db 2750 KVRKPMHESHTKLYQIDVMAHCLONTDVSLVSYNIQGVNDNRKPVFEADPYKAVLTEN 2809
OY 571 VPL-----YFVVASDKPRDQIGLGMSDGTGRLVTKEHNLINRD-----611
Db 2810 MPVGTSVIQTALIDKDPGRGQVSVRLSADPGSNV-----HELPAIDSESGMITTLOELD 2864
OY 612 -----GYOAGAEFLRS-----EDKKGVKL 630
Db 2865 CENCOTYHFHVAVADHGQITQLSSQALVOYSITDENNAPRFASEEYRGSVVENSEPEGL 2924
OY 631 YAT-KPLSHPLDQLRATLTCYQOE---VFGSTNGFD---LSTRLEHISRSIIQNG 681
Db 2925 VATLTADLADISBONRQVTCYITGDPGLGQFGISQVGDENMISRSKTLDRHTAK-----2979
OY 682 GMRRTYSLARKLRLKQAPPEPMODLPVDFVNGKPSQOEALLAGVAVHKYVADNLVNMRR 741
Db 2980 ----TLRLTASDGKFOA-SYVEIFVLVNDNSPOCSOLLYTKKHEDV-----3024
OY 742 GYRQYSLEVSSGLVSDAMMAIARAGISGVSPFGDNAVGSNRAHOMTGGIOACIYS--799
Db 3025 -FPGHFLKXASANDLDDTNAOI-----TYSL-----HGPB-AHEKRLDPHTELTTLT 3071
OY 800 -----DNFNHVPYRLRFPAGDQSTIRYAHDSLSPISDKGYLTGGQVLAAGTAEYNYE 852
Db 3072 ALDRERKDVFNLV---AKATDGGGRSCQADITLHVEDVNDNA-----PRF 3113
OY 853 FMKDLRLAVGCD--IGNAVYKGFNTDRTKIGAGVGRKASP---VGOVRVAVATGVKKEGN 907
Db 3114 FPHCAVAAYFNDTTVKTTPVAVVFARPDQGANAOVYSLDPSAGHPSIDATGTGVRLEK 3173
OY 908 PIRL 911
Db 3174 PLQV 3177

RESULT 10
US-09-815-242-12610
; Sequence 12610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl T.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
```

```
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12610
; LENGTH: 5795
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12610

Query Match 2.8%; Score 133; DB 10; Length 5795;
Best Local Similarity 16.7%; Pred. No. 7.5;
Matches 165; Conservative 144; Mismatches 348; Indels 332; Gaps 45;

OY 51 QAKAGNPVLLTPEQIOAR-----LNAAGLNAKQSOALDVNFDSDSPISRT---GEO 101
Db 792 RVRKPTPTTTTAEQLRGTALOKVPNVNTGISPLDPSALVYL-----APNQTNGSGSE 845
OY 102 SPLICIDMSVIEETPTPLSELEFAQESTEMGNPNDY---JPEYGEQ-----146
Db 846 ADQIPSGTIIATGTPDGVNHTIT-----IQPDYVVFIPV-GKQIRAVVYTNKYVA 897
OY 147 ---PNSEVVVPPTLEPERGLIKRLYARLFNDGVKKVPRLAKFYOSSQSGTSAIGSH 203
Db 898 SNMSNAVITLDDIPPT-----INNPVGINAKYRGDEVNFTMGVSDRH 941
OY 204 Q-KTEPYANIKAA-LEDTQESAMDUNGSIPLRQTLVLAARAVGYDDIDLSIIRNSIG 260
Db 942 SGLKSTITTLTPSGWTSNLTFSKSD--NKGSLATIGRVASMNA---FNSDITFRVSATD 994
OY 261 EVDYIHHDLGEPVYIDYRAVVEVRGEGADKAFTTVADDEVPLLCGFVHHKYEKKKMLE 320
Db 995 NVNNTTND-SQSKHVSINVKIS-----EDAHPIVLGN-----TEKVVVV 1033
OY 321 NASAE-----HGIFDGRWLDKRSVDVILPDNTADVSLIYDTGTQYRF 361
Db 1034 NPTAVSNDEKORITTAFMKNKQNTIRGYL-----ASTDPVTVVDNNGNVTLYLRDSSSTTL 1087
OY 362 DEVEFFTIDP--KTNOITTDPPDKLPVKRELLEOLLTVNMGEAVNLQAVRALNSDLIARY 419
Db 1088 DATVAVTTEPYVAKSEYOTANAKTAT-----VTAKGGSFNGIDK-----OY 1130
OY 420 FNMVTEIVPERQIONDOVSFEOSSSSRTEP-AQVDESTLEPIETVELTDGILMDS 478
Db 1131 FTLSNGOAI-----PQSTFTNITSDRITPNAQ-----1157
OY 479 PIESASNLIDOKILNVAAKARHLDMDD---RVLAINHDDGVNR---SILGRISDA 530
Db 1158 --EVSQNNAGTQLVHIVASNAHY--KDEDEFYISLKIYDVQKQPEDQRYRSTYDLTTDE 1214
OY 531 VSAVARAILPDESENEVIDPE-RTALANRKTPAVDY-----566
Db 1215 ISKVKQAFI--NARNDYITLAEGLISVTNTPNGANVSTIYVINKGRLTGSFASNLAMN 1272
OY 567 -----QSKKPELVYFVVASDKPRDQIGLGMSDGTGRLVTKEHNLINRDGYOAGAE 618
Db 1273 FLRWVNPQDVTYVWTVNVIKIANRPTDG--GLSWSD-----HMSLIYRYDATLTGTQ 1321
```

```

Qy 619 LRLSEDKKGYLVYTRKPLSLPLNDOLRAILGYQOEFG---HSTNGFDLSTLEHEIHSR 675
Db 1322 ITTMDILITMLKATTTVB---GLRNNITGNEKAAEAGGGPRNRTTGYSOANASSSDGQ--- 1375
Qy 676 SIQNGCGMNPRTYSRLRYLBDLKTQAPRPENMQDDPVDVFGKPSQSEALLAGVAVHKTVDN 735
Db 1376 -----KQYTL-----NGOYIQ-----TLD-1389
Qy 736 LVNPMRGYRORYSLSEVGSSSLVSDA---NMAIRAGISGVSEFGDAAGSNRAHOMTGCII 792
Db 1390 IINPSNGYGGOPPTNSTRANHSNATVYANNENPAAGACAFIT-DBHYKXSNSTHNASDAV 1448
Qy 793 QAGYIWSDFNEHWPY---RLRFFAGDQ---SIRGYAHDLSLPSDKLYLTGGQVLA----- 843
Db 1449 YKQALYTLPRGPQRQYVHLMQNQNGNTTDAINIIYFVPSDLVNPITISGANTNNQVSEETF 1508
Qy 844 -----VGTAEVYEFM-----KDLRLAVFGDIGNA 868
Db 1509 TNTITANDNEGVGSVTPPTSQLTGVDNHHQVSAATAPNATSATSKTINLATDTSGNT 1568
Qy 869 YDKGFTNDT-----KIGAGVGYRMAASPV 891
Db 1569 ATTSF-NVTVAKPLRDKRYVSTSTANPV 1596

```

```

RESULT: 11
US-09-815-242-13264
; Sequence 13264, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21, 078
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13264
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13264

```

Query Match	2.89	Score 132.5	DB 10	Length 495
Best Local Similarity	21.18	Pred. No. 0.18		
Matches 124	Conservative	79	Mismatches 197	Indels 189
			Gaps	33

Qy	148	NSEVVVPPTLEPEKGLIKR-LYARLFNDGVNKKVRLKAKFQSSO-SGQTSALISSHOK	205
2	SSKVVLT-TIFASGDLKKRKLYPSLF-----RLKRSNLSHPHVIQARR-	46	

QY	206	TEPANIKALEDITQESAMDLNSITRLKOTALVAARANGYDIDLSITRNSIGEVDI	265
Db	47	--PWS--KEYESVAVESIIDLADSTDOAEFA-----SHEYYO-----	81
QY	266	IHDGEPY-YIDYRAVEY-----REGADDKFTVADEVPILLIGDVFHNGKETKMLIE	320
Db	82	SHVDVDSCHYATLQOLQELNEKYQAEHNKLE--FLSMAPQFFETIAKHKSE--NIV-	135
QY	321	NASAEHGFEDRMLDVSVDVILLPNFADVSLIYDT--GTQY-----RFDEYVF	366
Db	136	-----DQGEER-----LIVKPPGTQYATANKLDELATFDEQI	172
QY	367	FTIDPKTNQLTTPDPKLPVKRELLEOLLTYNMGEAYMLQAVRALSNDLIATRYFNMYTE	426
Db	173	FRIDHY-----LGKEMIQSIFAVRFA--NLIFENVMNKKDFID-----NQ	210
QY	427	IVFPEREIOINDQVSEFOOSSSRPEPAQVDESLTPEVIEVELTDGLIMDISPESASN	486
Db	211	ITFAERLICEVERGGYDQSGALRP--MVQNHTL-----QLLSLAMP-KPASFYKDE	259
QY	487	LIQOKLNVAKAKHHLVMPDPRVLATINHDSQVRSILGRISDAVSAVARAILPDESENE	546
Db	260	IRAEKIKYF--KNLXH-PTDEBLKEHFIRGOYRS-GKI-DSMKYISYRSENE----	306
QY	547	VIDLPERITALNRKTPADVYQSKKVPYLVEFVASQKPRDGOIGLQMSGSDTCTRLATKFEH-	605
Db	307	-----VNPESTTFTTSGA-----FFVDSRFR-----GVFFFRTRGKRLLEKTHY	348
QY	606	NLINRDGYQAGAEURLSEDKKGKVLVATKPLSHPLDNOLRATLIGQOQVEGHSITNGDLS	665
Db	349	NIV-----FKQMSDIFGEELAPNI--LTIIYIQPTGEGSLS	381
QY	666	TRTLEHLSRSIT IONGCMNRKYSLRRLDKLKTQAPPTMODLFPVRYN	714
Db	382	LNGKQVGEFENIAPN-----SDYRFDATATASEPEYKELLIDVYDIN	423

```

RESULT 12
US-09-815-242-11134
; Sequence 11134, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815, 242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11134

```


QY 522 SI--LGRISDAVAVARAILPBDESEN-----EVIDLPERTALA 557
DB 1906 TVDYTOADSDKQNAKYQAIA--EAENVLKONSNOQVDAQOLNITLMAKQALNGDERVALA 1963
QY 558 NRKTPADVYSKKRPLVYFVASDPKPRDQIGLWGSPTGRLVTKFEHNLINRDGYQAGA 617
DB 1964 KTNCKHIDIDUAL-----NNQODGFKG-- 1987
QY 618 ELRLSEDKKVKLYATPLRSHPLN-----DQLRATLGYQOEYFGH---- 657
DB 1988 --RIDQ-----SHDLNQIOQIYDEAKALNRAMOL-----SOEISGNEGRT 2026
QY 658 --SINGNDLSTRLE-----HEISRSIIONGNNRTYSLRYRLDKLTKTOAPPEWQOLP 709
DB 2027 KGSITVYNADTQVQYVDEAVDAKQALDKSTGONLTAEOYIKLDAVTAAKKA----- 2080
QY 710 VDFVNGRP-----SOEALLGAVAVHKTVDNLVNPMPRGYRQRYSLVSSGLVSDANNAI 764
DB 2081 ---LNGERLNRKRSBALOR-----LDQTLHLNNAQRLAQIINNAETLNKASRAI 2129
QY 765 ARAGISGVYSFGDNAVGSNRHQMGTGGIOAGYIMSDNFNHYRLRFRFAGDOSIRGYAH 824
DB 2130 NRA-----TKLDNAMGA--VOQYIDEOHLGVISTNTYNADNMLK--ANYDNALANAAH 2179
QY 825 DLSLPSIDSKGYLTGGOVLAVGTAEYNEEFMKDLRLAVFGD--IGNAYDK--GFTN 875
DB 2180 E-----LDK---VOGNALAKAEADOLKONITIDAQNALNGDONLANAKKANAFAVN 2226
RESULT 18
US-09-915-242-5835
Sequence 5835, Application US/09815242
Patient No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815, 242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5835
LENGTH: 2434
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-915-242-5835
Query Match 2.8%; Score 131; DB 10; Length 2434;
Best Local Similarity 17.6%; Pred. No. 2.7;
Matches 181; Conservative 157; Mismatches 346; Indels 342; Gaps 46;

QY 29 LAAONNPANIIINHVAHDTALINOAKAGNPVLLPEQIOARLNAAGLNAPQOALDVVN 88
DB 1176 ISATINPEANPDITIQKASQVNSAKSA-----LNDE---KLAAKQKRAKSDIGSLTDLN 1227
QY 89 FDDOSPISRIEQQSPPLGIDMSVIEETPLSL-----BELFAOESTENGKINPDIIEBY 142
DB 1228 NAORTAANAEDQAPNLAAVTAARKKATSLNTAMGNLKHALLAEKDNTRKSVNYTQ----- 1282
QY 143 QGEOPNSEVVVPPTEPEPKGLIRKLYARLFNDGVNKNVPRLKAKFYQSQSESTSAIGSS 202
DB 1283 -ADQKQQA-----YDTAYTOAEAL-----TANAGSN 1308
QY 203 HOKTEPYANIKAALEDITQESAMDLNGSIPRLQYALYAARAVGY-----D 249
DB 1309 ANETO-----VOALNLQNLQ--AKNDLNGD--NKVAQAQKESAKRLALASYSNLNNQAOSTAATSO 1362
QY 250 ID-----LSTIRNSIEBVYIHDLCBPVYIDYRAVEVRGEGADKAFYTVADVEPLL 302
DB 1363 IDNATTVAGVTAAGVTANLNTAMGOL-----ONGINDQ--NTVKQOVNFT 1406
QY 303 IGDVFNHKKYEKRLNENASA-----EHGYFEDGRWLDSVDVILPD--MTADVSLIYDTGT 357
DB 1407 DAD---OGKKDAYTAAVTAAGCITLKAHQ---NMTKAQVEALNQYTTAKNALNGDANV 1460
QY 358 QYRFEDEVVFFITDPKTNQLTTPDKLPVYKRELLLEQL--LYVN-----MGEAY 402
DB 1461 RQA-----KSDAKANLGTLHLNNAQKQDLTQIEGATTVYNGVYKTKAQDLDGAMQ 1513
QY 403 NIOAVRALSNDLIATRYRNMVTEIYFPEREQIONDQVSFEQSSSRTPPAQVDESTLER 462
DB 1514 RLQSAIANKDQTKASE--NYIDAD--PKTKAFPMATITQAESYLNKHGANKDKQAOVQ 1568
QY 463 VTEIETLFDGILMDISPIEFSASNLIDOKLNL-----VAAKARHLIDMP 506
DB 1569 ALOSSTENALNGCANIORATEAIALDNLTLHTPQKTAALKOOVNAQAQVSGYTDK 1628
QY 507 DDRVLAINHDDGVNRSIL-----GRISDAVAVARAILPBDESENEV 547
DB 1629 NSATSLNNAMQDLQALADHDTIVASGNTYNAKSPKQAGYDTAAYNAKKIV--NGSPNVI 1686
QY 548 IDLPERTALANKTPADV-----YOSKVPILVYFASDKPRD--GQI----- 587
DB 1687 TNAADVTAATQYVYNNAAETGLNGDNLATATAKQOAKALROMTHLSDAQOSITGQIDSAFO 1746
QY 588 --GLWGSPTGTRL---VTKFEHNLINRDGYQAG----- 616
DB 1747 VTGQGVSKDNATNLNDNAMQNLNSTANKDKVAKSQPYVDADKQNAVYFATNAENIIN 1806
QY 617 -----AELRLSEDK--KGVKLYATK-----PLSHPLNQLRATLGT 650
DB 1807 ATSOPTLDPASAVTOAANQVSTKTKALNGAONLANKKQETTANINOLSHLNNAO----- 1859
QY 651 QOEYVGHSTNGFDLST---RTLEHEISRSI--IONGNNRTYSLRYRLDKLTKTOAPPET 704
DB 1860 KODLNTQVTNAPNISTVNOVKTKABLDQAMERLNGIDK-----DOVKQS----- 1906
QY 705 WODLPVDFVNGRPQOEFALLAGVAVHKTVDNLVNPMPRGYRQRYSLVSSGLVSDANNAI 764
DB 1907 -----VNFTDADPEKQTAANNV---TAENIINQANG-----TANQSQ 1943
QY 765 ARAGISGVYSF-----GDNAY--GSNRHQMGTGGIOAGYIMSDNFNHYRLRFRFAGGO 817
DB 1944 VEAALSTVYTTKQALNGDKKVTDAKNNAQOTLSTL-----DNLNN----- 1983
QY 818 SIRGYAHSLSLPSIDSKGYLTG--GOVLAVGTAEYNEEFMKDLRLAVFGDIGNAYDKGFTN 875
DB 1984 -----AOKGAVTGINQAHYAEVYQALIQTOELNTAM--GMLKNSLN---DK 2026
QY 876 DTKGA 881
DB 2027 DTTLS 2032


```

: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 1410
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 5229
: LENGTH: 1184
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
US-09-815-242-5229

Query Match      2.8%; Score 130.5; DB 10; Length 1184;
Best Local Similarity 18.4%; Pred. No. 0.97;
Matches 143; Conservative 131; Mismatches 287; Indels 215; Gaps 35;

QY 74 GLNAKQSOALDVNFDSPISRIQEOSPPLGLDMSVIEETPTLSLELFAOESTEMGI 133
DB 31 GPNGSGKSNITDAIKW-----VLGEQSAK-SLRGSKMEDIIIFSGAHRKQNAVEQL 82
QY 134 NPNDIPEYQEGQPSSEVVPPTL-----EPE-----KPELIRLRYAR 171
DB 83 RLDNNSKSLSYDE--NEVIVTRRLYRSGESESEYYINNDRARLKDIADLFDSGLGEAVSI 140
QY 172 LFNDGVNKPRLKAKFYOSOGSETSA--IGSSHQKTEPYANIKALEDIT--QESAMD, 227
DB 141 ISQGRVDEI--LNAKPIRROIIESAGVLKRRKRAESLNKLDQTEDNLTVEIDL 198
QY 228 NGSIPLRLQTLAVARAVGYDIDLSITRNSIGEVDT--IHDGEPV---YIDYRAVE 281
DB 199 EGRVEPLKEEAIAIK-----EYKTLHQMKHSDIYVTVHDIDQYTNDRNQDRLMD 250
QY 282 VRGEGADKATTVADDEVPL-----LIGGVFHGKRYETKKN 317
DB 251 LOGOANKE-----ADKORLSQOIQYKGRHOLDNYESLNYQLVKATEAFKETTGOIN 305
QY 318 LIENASAEHGYFDGRWLDRSVDT--LPDNTADVSLIYDTGTQYRFDEVFPTIDPKTNO 375
DB 306 VLEERKKKQSEFNARYEEOENLIELLENISNEISEADT-----YKSLSKQKE 355
QY 376 LTTDPDKLPVRELLEQLLTVNMGAEVNLQAVRALSNDLITRYFNMYNTELVPEPEROI 435
DB 356 LNA-----VIRELEEOLYVSDAEHDEKLEIK-----NEYTTLMS-----EQSDV 395
QY 436 QND-----QVSEFOSSSSTPEPAQVDESTLEPIYETVELTDGIL-----MDISPI 480
DB 396 NNDIRFLKHTIEENBAKKSRL-----DSRLVEYFEQLKIQOIKTKKEYOQTNKELSAV 451
QY 481 EFSASNLIDQKLNIVAAKARH--LY-----DMPDDR--LAINHDD-----GV----- 519
DB 452 DKEIKNIEKDLDTKKAKONEYEKLYQAYRYTEKKMKTIRIDSLAQOEEYVTFEFGVKHIL 511
QY 520 ---NRSLG-----RISDAVSAVARAI--LPRESENEVID-----LPERPAL 556
DB 512 KAKKKEKLGIGHGVAEIIIDVPSKLTQALGASLGHVIVDSEKDRGQAIQFLKERNIG 571
QY 557 ANRKTADVYOSKRVPL-----YVFVASDKPR-----DGQIG----- 588
DB 572 RATPLPLNLOSRYVANDIKSIKAEANGFISIAEAVKVAPEYNIIGNLGNITIVHL 631
QY 589 -----LOWGSDTGRVLTKFEHNLINNDGYQAGAEIRLSEDKKGVKLYATKPLSLPLNDQ 643
DB 632 KHANELARAIKYRTRIT--LEGDIVNPGSGMTGGARKSKISLSQKDEL--TMHQLEDY 689
QY 644 LRAATLGVOEVFGSTNGFDLSTRTLEHISRSIIIONGMNRTSLKRYRLDKLTKQ 699
DB 690 LROTSEFOQFKELIKISDOLSELVPEKSOAHNTLK-----EYVHFMELDRLTTO 741

RESULT 21
US-09-815-242-12125
: Sequence 12125 Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
```

```

: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA 011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 12125
: LENGTH: 1188
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
US-09-815-242-12125
```

```

Query Match      2.8%; Score 130.5; DB 10; Length 1188;
Best Local Similarity 18.4%; Pred. No. 0.98;
Matches 143; Conservative 131; Mismatches 287; Indels 215; Gaps 35;

QY 74 GLNAKQSOALDVNFDSPISRIQEOSPPLGLDMSVIEETPTLSLELFAOESTEMGI 133
DB 32 GPNGSGKSNITDAIKW-----VLGEQSAK-SLRGSKMEDIIIFSGAHRKQNAVEQL 83
QY 134 NPNDIPEYQEGQPSSEVVPPTL-----EPE-----KPELIRLRYAR 171
DB 84 RLDNNSKSLSYDE--NEVIVTRRLYRSGESEYYINNDRARLKDIADLFDSGLGEAVSI 141
QY 172 LFNDGVNKPRLKAKFYOSOGSETSA--IGSSHQKTEPYANIKALEDIT--QESAMD, 227
DB 142 ISQGRVDEI--LNAKPIRROIIESAGVLKRRKRAESLNKLDQTEDNLTVEIDL 199
QY 228 NGSIPLRLQTLAVARAVGYDIDLSITRNSIGEVDT--IHDGEPV---YIDYRAVE 281
DB 200 EGRVEPLKEEAIAIK-----EYKTLHQMKHSDIYVTVHDIDQYTNDRNQDRLMD 251
QY 282 VRGEGADKATTVADDEVPL-----LIGGVFHGKRYETKKN 317
DB 252 LOGOANKE-----ADKORLSQOIQYKGRHOLDNYESLNYQLVKATEAFKETTGOIN 306
QY 318 LIENASAEHGYFDGRWLDRSVDT--LPDNTADVSLIYDTGTQYRFDEVFPTIDPKTNO 375
DB 307 VLEERKKKQSEFNARYEEOENLIELLENISNEISEADT-----YKSLSKQKE 356
QY 376 LTTDPDKLPVRELLEQLLTVNMGAEVNLQAVRALSNDLITRYFNMYNTELVPEPEROI 435
DB 357 LNA-----VIRELEEOLYVSDAEHDEKLEIK-----NEYTTLMS-----EQSDV 396
QY 436 QND-----QVSEFOSSSSTPEPAQVDESTLEPIYETVELTDGIL-----MDISPI 480
DB 397 NNDIRFLKHTIEENBAKKSRL-----DSRLVEYFEQLKIQOIKTKKEYOQTNKELSAV 452
QY 481 EFSASNLIDQKLNIVAAKARH--LY-----DMPDDR--LAINHDD-----GV----- 519
DB 452 DKEIKNIEKDLDTKKAKONEYEKLYQAYRYTEKKMKTIRIDSLAQOEEYVTFEFGVKHIL 511
```



```

: TYPE: PRT
: ORGANISM: Helicobacter pylori
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: 85
: OTHER INFORMATION: xaa = Any Amino Acid
US-09-881-752a-308
```

```

Query Match      2.7%, Score 127.5; DB 10; Length 1797;
Best Local Similarity 19.2%; Pred. No. 3;
Matches 193; Conservative 147; Mismatches 354; Indels 309; Gaps 52;
```

```

QY 55 GNPVLLTPPEQIOARLNAGLNAGLNAPQSQALDVNFDDQSPISRIGESQSPPLGLDMSVIEE 114
DB 582 GAPEIIVRNEK-----RMLQEAVDVLFQNGRSTNAVKGANKRPL----- 620
QY 115 TTPPLSEELFAQESTEMGINPNDYIPEYQEQPNSEVVVPPTLEPEKPKLIRLYARLFN 174
DB 621 ---KSLSEIIKGKQGRFQNLGKRVDFSG---RSVIYVGPMLKMDCEGLPKNMALFLFK 674
QY 175 DGVNKVPRLAKAFYQSSQSGETSAIGSSHQKTEPYAN-IKALEDITQESAMDLNGSIPR 233
DB 675 -----PHLSKL---EERGVAFTTLKQAKRMIEQKSNEVWECLQETTEGYVLLNRA-PT 724
QY 234 LRQTA-----LVAAAVGY-----DID-----LSIIRNSIGE-----V 262
DB 725 LHKOSIQAFHFKLIDGKAIQLHPLVCSAFNADFGDDQMAVHVPLSQEALAECKVLMLSM 784
QY 263 DVIHDLGEPIYID-----YRAVEVRGEGADDAFTTVADEVPLL-IGDVFHNGKY 312
DB 785 NILPASGKAAVAIPQSDQVNLGLIYLSLEKSGVKGHEKLFSSVNEITTAIDTRELIDHAKI 844
QY 313 ET--KKNLIENASAEHGFDGRMLDRSDVILPD-----NTADVSLIYD----- 354
DB 845 RVLDQGNIIATSA-----GRMIIRKS---ILPFIPTDLMNRPKKKDDIGVLDVYHNV 894
QY 355 -----TGQYRFDEVVFETIDPKTNQTLTPDKLPVR--EL 389
DB 895 GCGIGTATFLDNLKTGLGRYATKAGISISMEDI-----ITPKDKQAMVEKAKVEYKKIQOQ 950
QY 390 LEQLLTVMGEAVYNLQAVRALSNDLIATRYFNMTETIYFPEREIQNDQVSFEQ----- 444
DB 951 YDQGLITDQ--ERY-----NKLIIDT--WTEVNDKMSKEMMTAIADQKEGFNSTYMAA 998
QY 445 SSSSRTERPAQV-----DESTLE-PVITVELTDCIILMDISPIESASNLID 490
DB 999 DSGARGSAAOIRQUSAMRGMLTKRPGSILTEPILISNFKEGLVLEYFNSTHGARGGLADT 1058
QY 491 KLNLVAA--KARHLVMDPDDRVLAINHDDGVNRSILGRISDAVSAVARAILPDESE--- 544
DB 1059 ALKTAANGYLTRKRLIDV--SQNVKVVSDDCGTHEGT--EITD-IAVGSLEILEERIGR 1114
QY 545 ---NEVIDLPRTALANKRTPADYVQSKV-----PLVYFVASDKPRD-----GOIG 588
DB 1115 VLLSDVIDIPITNEILLYADLTIDEAGAKKVEAGIKSITIRTPYCKAPKGCACACYGLN 1174
QY 589 LGWG-----SDGTSLVTKFEH-----NLINDGQOAGAE 618
DB 1175 LGEKMSYPCGAVGVVAAQSIGEPQTLLTRPHVGGTASNSODERETIVASKEGVRFYN 1234
QY 619 LRLEDKRGKVLATK-----PLSHPLNDQLRATLTGVOQDFVGHSTNGFGLSTRT 668
DB 1235 LRVTNKGKNIINIRRNASTLIVVEPKIKAPFDGELRIETIYEVEVVS----- 1282
QY 669 LEHLSIITONGMNRITYSLRRLDKLKTQAPPETMODLPYDVFYNGKPSQEFALL--AG 725
DB 1283 -----VANGDOEAKFVLR-RSDIYK---PSE-----LAGVGKIEGVVLPYASG 1323
QY 726 VAVHK--TVADNLVNPNGYQORYSLEWSSGLVSDANMATAR---AGISGYSGCDNAY 780
DB 1324 HKVYKGSIIADIT--QEGMNVNPNRIPYASSELVKD--NDPIADQVYAAKEGVIKY----Y 1375
QY 781 GSNRAH--OMTQGIQAGYIWSDFNHFVYRLRFFA-----GGDOSIRGY--AHDLSLSPISK 833
```

```

DB 1376 VLEANLERTHGIKKGDWSEK-----GLFAVIADNGREAAHYIARGSEILLIDN 1427
QY 834 GYLFGQVLAVGTAE-----YNYEFMKDLRLAVEGDIG 866
DB 1428 SEVSTNSVSKRPTTNFKTIATWDPYNTEPLIADFK-----GKVG 1466
```

```

RESULT 24
US-09-815-242-11410
: Sequence 11410, Application US/09815242
: Patent No. US20020061569A1
GENERAL INFORMATION:
: APPLICANT: Haseibeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: TITLE OF INVENTION: Identification of Essential Genes In
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT FILING DATE: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 11410
: LENGTH: 2890
: TYPE: PRT
: ORGANISM: Helicobacter pylori
US-09-815-242-11410

Query Match      2.7%, Score 127.5; DB 10; Length 2890;
Best Local Similarity 19.2%; Pred. No. 6.4;
Matches 193; Conservative 147; Mismatches 354; Indels 309; Gaps 52;
```

```

QY 55 GNPVLLTPPEQIOARLNAGLNAGLNAPQSQALDVNFDDQSPISRIGESQSPPLGLDMSVIEE 114
DB 1675 GAPEIIVRNEK-----RMLQEAVDVLFQNGRSTNAVKGANKRPL----- 1713
QY 115 TTPPLSEELFAQESTEMGINPNDYIPEYQEQPNSEVVVPPTLEPEKPKLIRLYARLFN 174
DB 1714 ---KSLSEIIKGKQGRFQNLGKRVDFSG---RSVIYVGPMLKMDCEGLPKNMALFLFK 1767
QY 175 DGVNKVPRLAKAFYQSSQSGETSAIGSSHQKTEPYAN-IKALEDITQESAMDLNGSIPR 233
DB 1768 -----PHLSKL---EERGVAFTTLKQAKRMIEQKSNEVWECLQETTEGYVLLNRA-PT 1817
QY 234 LRQTA-----LVAAAVGY-----DID-----LSIIRNSIGE-----V 262
DB 1818 LHKOSIQAFHFKLIDGKAIQLHPLVCSAFNADFGDDQMAVHVPLSQEALAECKVLMLSM 1877
QY 263 DVIHDLGEPIYID-----YRAVEVRGEGADDAFTTVADEVPLL-IGDVFHNGKY 312
DB 1878 NILPASGKAAVAIPQSDQVNLGLIYLSLEKSGVKGHEKLFSSVNEITTAIDTRELIDHAKI 1937
QY 313 ET--KKNLIENASAEHGFDGRMLDRSDVILPD-----NTADVSLIYD----- 354
```

```

Db 1938 RVLDDQGNITATSA-----GRMTIKS---ILPDEIPTDLNNRPMKKKKDIDVLVDYVHKV 1987
QY 355 -----TGQYRFDEVEFFETIDPKTNOLTTTPDPLPKR--EL 389
Db 1988 GGIGITATFLDNKLTGFRATKAGISIMEDI-----ITPDKOKMVEKAKVEVKIQOQ 2043
QY 390 LEOLLTVNMGDAVYLVQAVRALSNDLIATRYFMVNTIEIVPEREOIONDOVSFEQ-----444
Db 2044 YDQGLLTDQ-ERY-----NKIIDT--WTEVNDKMSKEMMTAIAQDKRGFSIYVMA 2091
QY 445 SSSSRPEAVY-----DESTLE-PIVETVELTGILMIDISPIEFSASNLIQD 490
Db 2092 DSGARSGAAOIROLSAMRGIMTKPDGSIETPIISNFKELINVEYFNSTGAKRGIA DT 2151
QY 491 KLNLVAA--KARHLYPDPRDLVLAINHDDGVNRSILGRISDAVASAVARAILPDSE-----544
Db 2152 ALKTANNGYLTKRLIDV--SONVKYVSDDCGTHGSI--ETID-IAVGSLEIPLERIFGR 2207
QY 545 ---NEVIDLPERTALANRKTTPADYQSKV-----PLYVFAVSDKPRD-----GOIG 588
Db 2208 VLEEDVIDPTNEILLVADTLIDEGAKKYVEAGIKSITITPTVCKAPKGVCAKCYGLN 2267
QY 589 LGMG-----SDTGRFLVTKFEH-----NLNRDGYQAGAE 618
Db 2268 LGGCKMSYPEAVGVVAAQSIGEPGTQLTLRTFHVGGTASRQDEREIVASKEGFVREYN 2327
QY 619 LRLESDKKGVKLYATK-----PLSHPLNDLRATLGQOEFGVSHNGDLSIRT 668
Db 2328 LRTTYNKEGNITIANRKNASILVVEPKIKAPFGDELIEVYEEVVS-----2375
QY 669 LEHEISRSITLONCGMNTYSILRYRLDKLTQAPETWODLVPVFNCKPSEQALL---AG 725
Db 2376 -----VKNDGQAKFVLR-RSDIVK---PSE-----LAGVGKIEBKVLLPVASG 2416
QY 726 VAVHK--TVADNLVNPBRGRQRYKSLSEVSGSLVSDANMAIAR---AGISGVYSFGONAY 780
Db 2417 HKVHKGGIADIT--OEGNNVPRRIPYASELLYKD--NDPIADVYAKKGVIKY---Y 2468
QY 781 GSNRAH--QMTGGIOAGYIGSDNFNHPYRLRFPA-----GGDOSIRGY-AHDSLSPI SDK 833
Db 2469 VLEANHLERTHGKIKKGDVSEK-----GLRAVIADNDGREGARHYIARGSEILIDN 2520
QY 834 GYLTGGVLAAGTAE-----YNYEPMKDLRLAVFGDIG 866
Db 2521 SEVSTNSVISKPTTNTFKTATWDPYNTPTIADFK-----GKVG 2559

RESULT 25
US-10-107-649-2
; Sequence 2, Application US/10107649
; Publication No. US20030044949A1
; GENERAL INFORMATION:
; APPLICANT: Behr, Regine
; APPLICANT: Sloma, Alan
; TITLE OF INVENTION: Polypeptides Having Gamma-Glutamyl
; FILE REFERENCE: 10157, 200-US
; CURRENT APPLICATION NUMBER: US/10/107,649
; PRIOR APPLICATION NUMBER: 60/279,374
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 604
; TYPE: PRT
; ORGANISM: Bacillus agaradhaerens
US-10-107-649-2

Query Match 2.7%: Score 127; DB 9; Length 604;
Best Local Similarity 21.1%: Pred. No. 0.61;
Matches 116: Conservative 87; Mismatches 226; Indels 122; Gaps 29;
```

```

QY 190 SSQGETAIGSSHOKTEPYANIKAALEDITQESAMD--NGSIPRLROTALVAARAVGY 247
Db 55 SENNGQTAEQNNGEARVYITGSSA--HPLAAVEGMDIENG-----TAIDAAVAVSF 106
QY 248 YDIDLSIR---NSIGEVYII-HDLGEPVY-IDYR-AVEVREGAGAD-----289
Db 107 M---LNVPEYGGIGGGVMLYHDPAGVSIYDYREAAPIGSGN--DDPTGCAVAPGV 161
QY 290 KAFITVAD-----EYPLLIGDYFHGHKIKETKKNL--IENASAEHYFD 330
Db 162 KGMDLIHNNHGELEPVEDYIAPAIEAETGFQVGDIFHOOTGNNAVYILMEEHEROLFPE 221
QY 331 GRMIDRSYVDILPMDTAVLSIYDGT-----QYRDEVVFFIIDPKTNOL 376
Db 222 GQALGVNDQVQEDLADTLRLIQENRSQGFYSGPIGLDLOQOFTFEDLASVFEQITE- 280
QY 377 TTDPDKLPVKRELLQOLLTVNMGDAVYLVQAVRALSNDLIATRYFMVNTIEIVPEREQIQ 436
Db 281 -----PVSAEVEGQIYVGGPSPSSGTVVQAL-----QVADQDLNDVFPD-EDLP 325
QY 437 NDQVSFEQSSS-----SRTEPAQVDESTLEPVETVELTGILMIDISPIEFSASNLIQD 490
Db 326 EDFESFGDSQHIYIHLINIEITKANSTKATYDSRLDT--LGPAPFDDIDHQAALTD DDIYIQ 383
QY 491 KLN-----LVAARARHLXMPDDRVLAINHDDGVNRSILGRISDAVASAVAR---ALLP 540
Db 384 LDDISFNEITPGDTSELDSPAEADSRHTTHFYIDKBERMASATSLSEFFGSGIYI 443
QY 541 DE---SENEVIDLPERTALANRKTTPADYQSKVPLYVFAVSDKPRDQIGLGMGSDTQTR 598
Db 444 DGEFINQMTFSDNPDSINRYEPGRKRTFVAVH-IF-----EEEGPVLMGSGPGRR 497
QY 599 L-VTKFE-----HNLINRDG---YQAGAEIRLSEDKKGVKLYATKPLSHPLNDLRATL 648
Db 498 IPAVFQTIMQYHGINDGDPMTLQEAIEAPRYNEEDV-IYLOEELPEVDSNELR-NM 555
QY 649 GYQOEVEFGHSF 659
Db 556 GY--SVVGHSS 564

RESULT 26
US-09-820-843A-16
; Sequence 16, Application US/09820843A
; Publication No. US2003039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PRO
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 1616
; TYPE: PRT
; ORGANISM: M. genitalium
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: cytochrome-c accessory protein
; NAME/KEY: misc feature
; OTHER INFORMATION: 9111046097
US-09-820-843A-16

Query Match 2.7%: Score 125.5; DB 9; Length 1616;
Best Local Similarity 19.5%: Pred. No. 3.6;
Matches 111: Conservative 94; Mismatches 180; Indels 183; Gaps 31;
```

```

QY 76 NAKPQOALDVVNFDDQSPISRIGQSPPLGLDMSVIEETTPLSLELFAQSESTEMG---132
Db 297 NTKSLASSL-----ENMLLENSSEDP-----VIEVKPRRNEVIFNPYTKLHFEX 344
```


;; PRIOR FILING DATE: 2000-11-28
;; NUMBER OF SEQ ID NOS: 24
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 10
;; LENGTH: 2310
;; TYPE: PRF
;; ORGANISM: Mus musculus
US-09-995-342-10

Query Match 2.6%; Score 124; DB 10; Length 2310;
Best Local Similarity 19.3%; Pred No. 8;
Matches 166; Conservative 119; Mismatches 307; Indels 270; Gaps 43;

OY 112 IEETPLSLEELFAOESTEMGINPNDYIPEYGEOPNSEVVPPTLEPEKPGI-----IK 166
DB 894 LEKTEPLT-EEMEDPEHPE-GMNDSPF-----ERELPGLVPGVCVK 932
OY 167 RLYARLFNGVKNVPRLLAKKFFIOS-----SOSCTSAIGSSHOKTEPYANIKALED 218
DB 933 NLVKNVEPSGRPAVDRLNTFYEMQITAFLGHNAGKTTTLSTI----- 975
OY 219 ITQESAMDNGSIPRLROTALVAVAGYDIDLSIIRNSIG---EVDYIHLGEPVYI 275
DB 976 -----LTGLLPSTSGVLLIGKDI---ETMLDVYKOSLGCMCPQHNLIFHHLTVAEHI 1024
OY 276 DYRAVEVKEGADKAFTTVADEPRLIGDY-FHNGKYETKKNL-----IENASAEHG 337
DB 1025 LEFYA-OLKGRSMEAOI-----EMEAMLEDYGLHKKRNEEAQDLSGCMOKKLSVALAFVG 1078
OY 328 YFDGKMLDSVYIILPDNTADVS-----LIYDTGYRDEVEVFTIIDPKTNOLTTDPDK 382
DB 1079 -----DSKV-VLDEPTSGVDYPSRRSIWDLILKYRSGRTIMS-----THHME 1122
OY 383 LPVKREILEQLITVMNGEAYNLQAVRALNSDLATRYFMVNVTEIYFPEREOIONOVSR 442
DB 1123 ADL-----LGRIRAIISGRKICSGTPLYLNCSTGTGYLTVR-----KKNIQSGGCG 1173
OY 443 EOSSSSRTE-----PAOVDESTLEPVI-----ETVELTGDILMIDISPIEFSSANLI--- 488
DB 1174 EGVCSCTSGKSTFCTPRTYRDEITEBQVLDQVQELMDVLYHNHPEAKLVCEIQELIFLL 1233
OY 489 -----QOKNLVAAKARHLVMDPDYV-LAINMDGVNRSILGRISDA 530
DB 1234 PNKNFKORAYASLFRLEETLADLISGFSISDTPLEELFLKYTEAGAGSMFVG---GA 1290
OY 531 VSAVARAIL-----PDESENEVIDLPERTALANKRTPADYVQSKVPLVYFVASDKPRD 584
DB 1291 QOKRQOAGLRHCSAPTEKLRQYAOAPHTC-----SPGOVDPKGP-----SPEPED 1338
OY 585 GOIGLGWGSDTGTR-----LVTKFEHNLINRDGYOAGAEILRLSEDKGKGLYATKP 635
DB 1339 P-----GVPEFTGARLILQHVQALVYKRFHHTIRSKDFVQAQIYLPATFYFLALMLSLIYP 1394
OY 636 LSHPLNDOLARILGYOQEVFGHSTNGFDLSTRTLEH-ELISRTION--GGMNRTYSLRYR 692
DB 1395 ---PGEERPALTL--HPWMYGHQYTFESSMDEPNNELVLAUVILLRPFGNRC----- 1443
OY 693 LDKLTQAPPE-----TWODLPYDFVNGKPSQELI---AGVAVHKVY----- 732
DB 1444 ---LKEEMLPEIYPCINATISMKTPSV-----SPNITHLFKOKWMTAHPSPCKSCSTREKL 1495
OY 733 -----ADNLVPMRGYRQYSLVSGSSGLVSD---ANMAIARAGISGVYSGDNA 779
DB 1496 TMLPECPBGAGGLPPQRFQSTEVLDLTNNINISDYLVKTYPALRRSLKSKFWNNEOR 1555
OY 780 YGSNRAHOMTGTIOAGYISDNFNHVPYRLRFPAGGDSIRGYAHDSLSPISDK---GYL 836
DB 1536 Y-----GGISIG-----GKLP-----AIPISGEALVGL 1579
OY 837 TG-GOVLAVGTAEVNYEEMKOL 857
DB 1580 SGLGOMMNVSGGPVTRASKEM 1601

RESULT 29
US-09-801-368-364
; Sequence 364; Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sheeraman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 364
; LENGTH: 1250
; TYPE: PRF
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-364

Query Match 2.6%; Score 123.5; DB 10; Length 1250;
Best Local Similarity 20.0%; Pred. No. 3.4;
Matches 209; Conservative 138; Mismatches 363; Indels 337; Gaps 58;

OY 33 NNP-----ANINNVPAHDIAINQAKGNPVLTPPQIQARLNA--AGLNAPK 79
DB 140 NNGGNSHRRKTSQSSITYGHSRRHSGLNKAIRA-----AAEQAKRISGEGAGYVK- 192
OY 80 QSOALDVNFDQSPISRICEOS-----PPLGLDMVIEETPLSL- 120
DB 193 -----IDSVQADSS--NSTIEQSDFKFPPPNAAHQHRRATSNLSPSEKFPNNSHCND 246
OY 121 EELFAQEST--EMGINPNDYIP-----EYQGEOPNSEVVPV-----TL 157
DB 247 DEFIATSTFHRSKTRNNEYSFGINSWMRQSQOPOQ--LSPFRRHGSNSRDYNSPNTL 304
OY 158 EPEKPGILIRLYA-RLEFNDGVKNVPRLLAKKFFYQSSQSGETSAIGSSHOKTEPYANIKAL 216
DB 305 EP--PAIFQDGHKHRASNSVHSF-----SSQGNNGN--GGRKSLFAPYLP-QANI 350
OY 217 EDTQESAMDNGSIPRL-----RQIALVAVAGYDIDLSIT-----RNSIGVNDYIHD 268
DB 351 PELIQGR--LVAGILRVKKNRSDAMVSTD--GALDADYITGSGDRNALBGLV--- 403
OY 269 LGEPRYIDYRAVEVKEGADKAFTTVADEPRLIGDYFHHGKYETKKNLIENASAEHG 328
DB 404 -----AVEL-----LVVDVWESKKKEKKEKKRKAASQOHD 435
OY 329 F-----DGKMLDSVYIILPDN-----TADVSLIYDTGYRDEVEVFTIIDPKTNOLTT 378
DB 436 IPLNSSDDYHNDAVSYAATSNNEFLSSPSSSDLSKDDLSYR-----KRSTINN 485
OY 379 DPDKL--PYK-----RELLEQLITVMNGEAYNL--QAVRALNSDLATRYFMVNVTEIYF 429
DB 486 DSDSLSSPTKSGVRRRSSLKQPTQKKNDVEVEGOSLLIVEEELNDKYPKYAGHYV- 544
OY 430 PEREQIONDOVS-----FEQSSSSRTEPAOVDESTLEPVEYVELND-----GILM 475

```
Db 545 AVLDIRIGOLFSGTGLLRPSQOANSNNKRPQ-----SPKIAMFKPTDKKVLPIAIP 598
QY 476 DISIER-SASNLIOQKLNLYAAKARHLYDMPDRVLAINHDDGVNSILGRISDAVASAV 534
Db 599 ELAKRDEVNADKSEKLFVASIKRWPITSL-----HPGILVSELGDIDHPDTEI 649
QY 535 ARATLPDES--ENVVID-----LP-BRTALANKTRPADVYQSKKVLVYVA 578
Db 650 -DSLRRNNFLSNRYLDOOKNPKPKPSEFQPLPLAESLEYRNFTDINEYN-----IFAI 703
QY 579 SDRPRDQOIGLWGS-----DTGRLVYKFEHNLINRQGVQAGAEIRLSED 624
Db 704 SE-----LGWVSEFALHYRNNGNGLTELGCHVVDVTSI-----IEGSSVDRRAR 748
QY 625 KKGVKATKPLSHPL-----NDOLRATLGYQOEVEFGHSTNGFDLSTRTL-----EHE 672
Db 749 KRSSAVMPKLVNLLPQSFNDELSTLAPGKESATL---SVYTTLDSSYLRKRSTWVGEST 805
QY 673 ISRSIIIONGCMNRTSYRLYRDLKLTQAPR---ETWODLPVDFVNGK--PQOEALLAGVA 727
Db 806 ISPSNI-----LSLEQDEKLSTGSPSTSYLSTVOEIANSFYARRINDEPATILLPTLS 857
QY 728 VHKTVAD-----NLVPMRGY-----RORYSLEVSSGLVSDANMAIAR 766
Db 858 LLESIDDEKVKVDNLTIDRTLGFVYVINEIKRKVNSTVAEKITYTKLGLDALLRQMOPIAT 917
QY 767 AGIS---GVYSFGNAXGSNRAHOMTGIO-----AGYI 797
Db 918 KMASFRKKIONFEGYN-EDTNTADELIGVLYKIKDDVAVGIEILLFKTMPRARYFIACKV 976
QY 798 WSDFNHVPYRLRFAGAGDSIRGVA---HDSL-SPISDKGYLGGOVLAAGVTAEYWE 852
Db 977 DPQGYHAYLNLPTHTFTAPMRKRYADHVYRQAKAYIHDPPTYEDMALAI-TSEY-CN 1034
QY 853 FMKD-----LRLAVFGDIGN 867
Db 1035 FKDKCAYAOEOAIHLLCKTINDMGN 1061

RESULT 30
US-09-738-626-6999
; Sequence 6999, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738.626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6999
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6999

Query Match 2.6%; Score 121.5; DB 9; Length 852;
```

```
Best Local Similarity 21.0%; Pred. No. 2.6;
Matches 193; Conservative 113; Mismatches 318; Indels 295; Gaps 51;

QY 1 MSKPVLFANRSEFMPALAAVPLMTSQALAOQNNPAN-IINHPADHTAINQAKGNPV 59
Db 45 VAAPPLMATGVDPKEILLAEAKLVASYPKASGAMNANPNFNDAIATFAQELAGE--- 101
QY 60 LITPEQIARLNAAGLNAPQSOALDV-----NFD----- 90
Db 102 -LGDEYSTEVLLAGI-ARQSDAADLLTNKGATYDAIKAPPSVRSQRYTTODPEQGF 159
QY 91 -----DQSPISRIQSGSPPLGDMVIEETPTLSLEELFAQSTENGIMPNDYIPY 142
Db 160 QALEKYSTDLTKLAEKGKIDPYIGRQDEIRRVQYLSR-----TKNN--PVL 205
QY 143 OGEOPNSEVVPPTLEPERKGLIKRLVYARLFENDGVNKPVR-LKAFYQSSQSGETSAIGS 201
Db 206 IGEPGVKTAI-----VEGLARRIVA-----GDVPESLKGKTLISLDIG--SMVAG 249
QY 202 SHOKTEPYANIKAALEDI-----TOESAMDNGSIPRL---RQ 236
Db 250 AKYRGEFEERLKAVIDEIKANGAEVYTFIDELHTTVGAGASGSEAMDGNMKPLLANGE 309
QY 237 TALVAVARAVGYDDIDLSIIRNSIGEVDYIHDGEPVYIDYRAVEVRGEGADDKAFTYA 296
Db 310 LRLVQATTLINEY-----RKTI-EKDAALERRPQOYV-----GE-----PYVE 346
QY 297 DEVPLIG---DVFHGKYETKKNLIENASAEHGYEDGAWL-DRSYDVILPDNTAD-V 349
Db 347 DAIGILRLKEREYV-HHGVRIODSALVAEELSRYITSRFLPKAIDLV--DEAASRL 403
QY 350 SLIYDTGYQYRDEY--VEFTIDPKTNOLITDPD-----KLPVKELLEQLIT 395
Db 404 RMEIDSSPO-EIDELERIVRLIEEMALSKESDASKERLEKRLSELADEKEREKSEIKA 462
QY 396 VNMGEAVYMLQAVRALSNDLIATRYENMNTETV-----FPEREQIONDQ 439
Db 463 RMONKETAIDVREKKELEALR-----SESDAENDGNVGRVAELRYGRIPLEK-----Q 514
QY 440 VSFQSSSSRTPEACV-DESTLEPYIETVELTGDILMDISPIESASNLIOQKLNLYAAK 498
Db 515 IEDASKVEVENMAMLTREEVTPDTIADVVASMTGI-----PAGKMMQ-----GE 558
QY 499 ARHLYDMPDDVLAINHDDGVNSILGRISDAVASVARAILPDESENYIDLPERTALAN 558
Db 559 TEKLINM--ERYLG-----NR-VVGOL-EAVTAVSAV--RRSRAGVAD-----PN 598
QY 559 RKTADVYQSKKVPPLYVAVASDKPRDQOIGLWGSDTGTRLVTKFEHNLINRD----- 611
Db 599 RPTGS-----FRLP-----GPTGVG-----KTELAKAAVEFLPFDDRAMIRID 636
QY 612 -----GYQAGAEIRLSEDKK--GVKLYATKPLSHPLINDOLRATL 648
Db 637 MSEYGEKHSVARLVGAPPGYVGDGGLTEAVRRPPTYVILFDEVEKAHP--DVFDITLL 694
QY 649 GYQOEVEFGHSTNGFDLSRTLEHE---ISRSIIIONGCMNRTSYRLYRDLKLTQAPPET 704
Db 695 QYDDE--GRLLTG---QGRIVDFRNTIILISNLGAGG-----TREQMDAVKKAAPKEEF 744
QY 705 WQDLPVDFVNGKPSQOALLAGVAHV-KTVADNLVPMRGYRORYSLLEVSSGLVSDANMA 763
Db 745 VNRLDDVVIYFDRLSPQEGTISIVDIQIKQLTDRLAG-----RLNLRVSDS-----AKAW 793
QY 764 IARAGISGVISFGDNAYGS 782
Db 794 LAERGY-----DPAYGA 805

RESULT 31
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
```



```

; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match      2.6%; Score 121; DB 10; Length 1430;
Best Local Similarity 18.8%; Pred. No. 6.3;
Matches 216; Conservative 160; Mismatches 396; Indels 378; Gaps 59;

Qy 25 TSOALAOQNPNANIINHPAHDATINAKA-----GNPVLITP-----BOIOA 68
Db 124 TSEERTKQSEEA-----QTASSOALQAKAKELTKQKQTAQENKKNVDLAIPNVQIOG 179
Qy 69 RLNAAGLAKPQSQ-ALDYN-----FD-DQSPISRGESPLGLDMSVIEETP----- 117
Db 180 KYYTIGSDGQPKRKALTYNNKVLXFDKNTGALTDSQYQFGKL-TKLNNDYTPHNQIV 238
Qy 118 ----LSLELFQOESENGINDYIPEYQOEPNSEVVPPLLE-----PEKGLIKRLA 170
Db 239 NFENISLETIDMYTADSYRKRKDLTKNGKTWTASESDLRPLKMSWWDKQTOI--ATL 296
Qy 171 RLFDND-----GVNKVP-----RLKAFYQSSQSCFSAIGSSHOKTEPY 209
Db 297 NYMNOQGLGTGENYTAADSSQESLNLAAQTVQYKIEKISQIQOTQMLRIINSFYKTP- 355
Qy 210 ANIKAALEDITQESAMDINGSTIPRLROTALVAAA--VGYYDIDLSIR----- 256
Db 356 -NMNSQTESDTSAGEKD-----HLOGGALLYSNSDKTYANSVDYRLNRTPTSQTKPK 408
Qy 257 ----NSIGEVDYIIH---DLGEPY-----YIYIRAY-----EVRGGADDKATF 293
Db 409 YFEDNSSGCTPFLANDINDNSNPVQAOQLMWLHYLMNNGSITVANDPEANFGVRAVD 468
Qy 294 TVADEVPPLLDVVF--HHGKYETKKNLIENASAEHGYFDG--RWLDRSVDVLP--DNFAD 348
Db 469 NVNADLLQIASDYKKAHGYGVKSEKNALNHLSILEAMSDNDQYKDKTGADLPIDNKR 528
Qy 349 VSLIY-----DTGTQ 358
Db 529 LSLIYALTRPLEKDAKSNKREISGLEPIYTNLSNRSAGKNSERMANIIFIRADSEVQ 588
Qy 359 YRPDEVVFTIDPKTNOULTDPDKLPV-----KRELEDLVLTNMGCAVNLQAVRAL 410
Db 589 TVAKITIAQINPKTGDGLFTLDELKQAFKIYNEDMROAKKKYQSNIPYAVALM-----L 644
Qy 411 SNDLIATP-----YFNMYNT-----EIVFPEREQIO 436
Db 645 SNKSITPFLLYGDMYSDGQYMATKSPYYDAIDTLKARIKAAAGQDMKITIYVGDSSH 704
Qy 437 ND-----QVSFEQSSSSRTEPACQVDESTLEPIETVELDGLIMDISP 479
```

```

Db 705 MDMDYTGVLTSVRYGTGANEDTQGEATKQGMAYITSN-NPSLK-LANQDKVIYVMGA 762
Qy 480 IERSASN-----LIDDKLNLVA-----AKARHLXMPDDR-VLAIHHDGVNRSITIGRISDAY 531
Db 763 AHNQOEYRPLLTTKDGLSYTSDAAKSLYRKTNDGELVFPASD-----IGYLNPOV 817
Qy 532 SAVARAILP-DESENEVIDPERTALANR-KTPADYQSKKV-----PLYVFA 578
Db 818 SGYLAWVPYVQASDNDQV-----RVASNKANATGOYIESSALDSQILYGFSPFQDFVT 873
Qy 579 SDK-----PRDQIGLNG-----SDTGTFLVTFEHNLIINRDGYQAGAEI 619
Db 874 KDSDYTNKKIAQNVQLFKSGWVTSFEMAPQYVSSSDGSFDSITIQNGYAFEDRY----DL 929
Qy 620 RUSED-----KKGVKXVATKRLSPHNDQRLATIGYQOEYVGHSTNGF 662
Db 930 AMSKNNKYGSOQDMINAVKALHKSIGIOWADW-----VPQIYNLPG--KEV-TATRVN 981
Qy 663 DLSTRTLEHISRSII-----ONGMNNRYSLRYR--LDKLTQAPPETWODLPVDFVN 714
Db 982 DYGEYRKDSIKKTLVLAANTKSNG---KDYQAKYGGAFLESLAKYPSIFNR---TOISN 1035
Qy 715 GK-----PSQEL-----LAGVAVHKTYADNLVN--PMRGYR---QRYSLVGS 753
Db 1036 GKSIDSEKITAWKAKYFNGTNIILGRGVYVLKDNASDKYFEIKGNQTYLPRQMTNKEAS 1095
Qy 754 SGLVSDAN--MAIARAGISGVTSFGDNATGNSRAHQTGCIQAGIYWSDNFNVPLRPF 811
Db 1096 TGEVNDNGMTEYTSYGQAKNSFQDAGKN-----WYFDMNGHMVY----- 1138
Qy 812 FAGQDSIGRYAHDLSPISDKGYLTGGOVLAVGTAEVNEFEKDLRLAVFGDIGNAYDK 871
Db 1139 ----GLOQLNGEVOYFLS-----NGVOLRESFLENADGSKNT-----FGHLNRYSN 1181
Qy 872 G---FTNDTK 878
Db 1182 GYYSFNDNSK 1191

RESULT 32
US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. US2002015568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHOUTZEN, LOBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-2

Query Match      2.6%; Score 121; DB 9; Length 1781;
Best Local Similarity 18.0%; Pred. No. 8.8;
Matches 173; Conservative 119; Mismatches 326; Indels 344; Gaps 44;

Qy 133 INP-----NDYIPEYQOEPNSEVVPPLLEPEKPGILKRLYARLFNDGVNKPRLAKFYQ 189
Db 668 VNPALSNSSMSGFQG-----ITLTPVYKNEVNUQLVHR-----FSN 703
Qy 190 SSQSGETSAIGSSHOKTEPYANIKAALEDITQESAMDINGSTIPRLROTALVAAAVGYTD 249
```

```
Db 704 DVKGBEN-----YDFWSELMPPV-KDSFQKNGP---LKQGLGTINGQOQY- 747
QY 250 IDLSIINSICE--VDYIHLGEPVYIDYRAVEVREGADDAKFTYVADVPPLIGDFV 307
Db 748 -----LPTTGQPKRNLLOSGNMWTFD-----SDTGVNTALELQRAKGYVS 791
QY 308 HNGYEFKRLIENASABHGYFDGRWLDORSVDYLLPONTADVSLYDTGTQYREDEVV-- 365
Db 792 SNEQY-----RNGNMAYSYD-----DKSIEVNGYLLAD-----TWYRPKQLIKD 831
QY 366 --FTIDPKT-----NLT-----TDPDKLPVKR 387
Db 832 GTTWTDSKETMDRPIILMMWPNLTLOAYLYLNMKOHGNLPLSPFPFNADAPALNHY 891
QY 388 ELLEBOLLTVNGEAYNLQAVRALSNDLIATRYFEMVNTETVEFPEREQIONDOVSFEQSS 447
Db 892 EIVQONIEKRISETGNTDMLRTLHDFVTNNPMKNKSENVNFGIOFGGFLKYENS DL 951
QY 448 S-----RTPEAOVDESTL-----EPVETVELT----- 470
Db 952 TPVANSOYRLGMPRIKQOTYRGOEFLLANDIDNSNPVVOAQOLMWLYLLNFGTITA 1011
QY 471 -----DGLIMDI-SPIEFASNLIDOKLNLV-----AAKARHLYDMPDRVLAINH- 515
Db 1012 NNDQANFDSYRVADPMDIDADMIAADYFNAAYGMSDAVSNKHINILED-----WNHA 1066
QY 516 -----DDGVNSILGRISDA-----VSAAVARAILPDESENEVIDPERTA 555
Db 1067 DPEYFNKIGNPQULMDTITKNSLNHGLSDATNRMGLDAIYHQSADRENNST----- 1118
QY 556 LANKKTADVQSKKVPYLVFV--ASDKPRGSOGLGMSDGTGRLVYKFEHNLINRQGY 614
Db 1119 -----ENVYIPNTSFVRAHDNNSODQI-----ONALPDYGVKDYHTTFEEDQ 1161
QY 615 AGAELRLSEDKKGVKLY--ATKPLSHPLNDQLRATLG--YQOEYF-----GHSTNGFD 663
Db 1162 KGIDAYIODONSJYKRYNLNIPASVAILLTNKTOTIPRYVGYDGTGQGWMEHQTRYD 1221
QY 664 LSTRLEHEL-----SRSLIONGCMNRTY--SLRY-----RLDKIKTQAPPETW 705
Db 1222 TLTNMLKSRKYYVAGGOSMOTMSYGNNMILTSYRGKAMTATDTGTDERTGIGVVV 1281
QY 706 QDLPEVDENKGPSEALLAGVAVHK-----TVADNLVN-----PMRGYRORYSL 749
Db 1282 SNRP-NIKLGVNDKVVYAHMG-AAHKNOQYRAAVLTITDGYINTISDQAGAPYAMDENGLD 1339
QY 750 EVGSSGLV-----SDANNAI-----ARAGISGVYSFGDNAY 780
Db 1340 YLSSHNLVYVNGKEADPAVOGYANPDVSGYLAVVVPVGCASDNODAPRSTKNSGSAY 1399
QY 781 GSNRAHQMTGICAGYIWSQNFNHPYR-----LRFAGDQDSIRGAHDSLS 828
Db 1400 RTMAAFDSNVIFEA---FSNFYTTPTKESERANVRIAQNADEFFAS-----LGFSTSEMA 1450
QY 829 P-----ISDKCYLITGGGYLAAGTAENYEFMKDLRLAVFCDIGNAYDKFTNDRTICAGV 884
Db 1451 PQYNSKSDRFLD-----STIDNGYAF-----DRYDLMSEBPNKYGIDED 1491
QY 885 VR 886
Db 1492 LR 1493
```

RESULT 33
US-10-060-036-71

; Sequence 71, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.

```
; APPLICANT: Hepler, William T.  
; APPLICANT: Jiang, Yugu  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER  
; FILE REFERENCE: 210121.566  
; CURRENT APPLICATION NUMBER: US/10/060.036  
; CURRENT FILING DATE: 2002-01-30  
; NUMBER OF SEQ ID NOS: 4560  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 71  
; LENGTH: 1192  
; TYPE: PR  
; ORGANISM: Homo sapiens  
US-10-060-036-71
```

Query Match 2.5%; Score 120.5; DB 9; Length 1192;
Best Local Similarity 18.2%; Pred. No. 5.2;
Matches 120; Conservative 102; Mismatches 245; Indels 193; Gaps 30;

```
QY 2 SKPVLFANRSEFVALAAYLPLMTSQALAOQNNPANTINHYPAHDTAINQAKGNPVL 61  
Db 449 STPGIKDRSGAYITCAFPNPAAT-ESIA-----TNIFPLGDPYSEKTKTEKKIEE 499  
QY 62 TPEDIQARLNAGINAKP-----OSQALDVYVNPDDQSPISIRIGESPLGIDMSYIETT 116  
Db 500 KKAQIVTEKNTSTISNPFVLAADSETDYVTTDLNLYTEVEYVANNPEGLTPDLVQAC 559  
QY 117 PLSEEL-----FAQES-----TEMGINPDYIPEYQGO 147  
Db 560 ESEINEVYGTKIAEYTKMDLVQTSFVMOESLYPAQOLCPFESEATPSPLDVIEMAP 619  
QY 148 -NSEY-----VPPYLEPEKPGIKRIYARLNDGVKKYRLAKRFQSSQSGTSAIG 200  
Db 620 LNSAVPSGASVIOPSSPLEASSVN--YESIKHEPENPP-----YEEAMSVLKKVS 671  
QY 201 SSHQTEPYANIKALIEDITQ---ESAMDLNGSIFRLQTAALVAAVGYDI-DLSIIR 256  
Db 672 GIKKEIKEPENINMALQETEARYSISICDL-----IKETLSAPADPDSYSMAKVE 725  
QY 257 NSICEVDYIIDL---GEPIYIDYRAVEVREGADDAKFTV---ADEVPILLIGDVFHGG 310  
Db 726 QPVPDHSLEYDSSPDSEPVLD-----FSDSIPVPQKQDFTVLYKESLRET 774  
QY 311 KYET-----KNLEMSAHEG--YFDS--RWLDRSVVILPDMTADVSLYDGTQYRF 361  
Db 775 SFESMIETENKISALPPEBGKPYLESFKLSLNDTKTLLPDEVSTLS----- 823  
QY 362 DEVEFTIDPKTNQITDPPDKLPVKRELLQILTVNMGEAYNLQAVRALSNDLIATRYFN 421  
Db 824 -----KKEKIPQ---MELSTA-----YISND----- 843  
QY 422 MVNTEIYFPEREQIONDOVSFEQSSSKRTEPAQVDESTLEPIYETVELTGLIMDISPIE 481  
Db 844 ---DLFISKEAQIRTE--FESDSS-----PIELIDFEPPLISKT 879  
QY 482 FSASNLIQDKNLVAAARHLYDMPDDR---VLAINHDGVNSILGRISDAY----- 531  
Db 880 DSFKLAREYITDLVSKSEIANNPDGAGSLPCTELPHDSL-KNIQPKVEKISFSDDF 938  
QY 532 ---SAVARAIL--PD-----ESENEVIDLP-RTALANKRTPADVQSKKVPYLVFVA 578  
Db 939 SKNGSATSKVLLPDPVALATQAEISIVPKVLYVKEAEKLPDSREKEDRSPSAIFSA 998
```

RESULT 34
US-09-758-140-6

; Sequence 6, Application US/09758140
; Patent No. US20020012965A1
; GENERAL INFORMATION:
; APPLICANT: Strittmatter, Stephen M.
; TITLE OF INVENTION: No. US20020012965A10 Receptor-Mediated Blockade of Axonal Gro
; FILE REFERENCE: 44574-5073-US
; CURRENT APPLICATION NUMBER: US/09/758.140

```

: CURRENT FILING DATE: 2001-01-12
: PRIOR APPLICATION NUMBER: US 60/175,707
: PRIOR FILING DATE: 2000-01-12
: PRIOR APPLICATION NUMBER: US 60/207,366
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 60/236,378
: PRIOR FILING DATE: 2000-09-29
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 6
: LENGTH: 1192
:
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-758-140-6

```

Query Match	2.5%	Score 120.5	DB 10	Length 1192
Best Local Similarity	18.2%	Pred. No. 5.2		
Matches 120	Conservative 102	Mismatches 245	Indels 193	Gaps 30

```

QY      2 SKPYLFAHSEFVALAAYLPLMTSOALAQOONPANIINHVAHDPTAINQAKAGNPVLL 61
Db      449 STPEIKRSGAYITICAFENPAAT-ESTA-----TIFELLDPTISENKTDEKKIEE 499
QY      62 TPEOIQARLNAAGLNAP-----QSOALDVNFDDQSPISRIGEOSPPLGLDMVIEETT 116
Db      500 KKAOLTEKNTSTKTSNPFVLAAGQSESDIYITDNLTVTEEVANMEBGLTDPDLQVENC 559
QY      117 PLSLLEL-----FAOES-----TEMGINPDYIPEYQGBOP 147
Db      560 ESELENYVGTRIAYETKMDLVQTSSEVMQESLYPAQLCPSESEBATSPLYLPDIYMEAP 619
QY      148 -NSEY-----VVPETLEPEKRGCLKRLIARLFNGCVKAVPRLAKFTQSSQSGTSAIG 200
Db      620 LNSAVPSGASVNIQOSSSPLASSVN--YESIKHEPNPP-----YEAMSVSLKRV5 671
QY      201 SSHOKTERYANIKALEDITO---ESAMDLNGSIRLROTALVAAVAYGYDI-DLSIIR 256
Db      672 GIKBEIKEPENINALOETEAPYISACDL-----IKETLSAEPAPDFSDYSSEMAKVE 725
QY      257 NSIEBVYI IHDL--GEPPYIDYRAVEVRGEGADDKAFTTV--ADEVPLLIGDVFHHG 310
Db      726 QPVDHSELVEDSSPDSESPVDL-----FSDSDIPVPOKQDPTVNLKESLLET 774
QY      311 KYEP-----KNLLENMSAENG--YFDG--KMLDRSVUYILPNDTADVSLIYDTGTQYRF 361
Db      775 SFESMIEYENKEKLSALPEEGGKPYLESFKLSLDNTKDTLLPDEVSITLS----- 823
QY      362 DEVEFTIDPRTNQUTDPODKLPVKRELLEOLLVYNNGEATINLOAVRLASNDLIATRFN 421
Db      824 -----KKEKTIPLQ-----WEEELSTA-----VYSND----- 843
QY      422 MVNTEIVPEREQIONDQVSEFOGSSRTPEAQVDESTLEPIVEFVELTDGILMDISPIE 481
Db      844 ---DLFISKAQGRRETE-FTSDSS-----PIELIDEPPLIISKT 879
QY      482 FSNASNLIQDKLNLVAAKARHLVMPDDR---VLAINHDDGVNRSILGRISDAV----- 531
Db      880 DSEFKLAREYTDLEVSXKSELIANAPDGAGSLPCTELPHDLSL--KNIQKVEKISFSDPF 938
QY      532 -----SAVARAIL--PD-----ESENEVIDLPE-RTALANRKTTRADYQSKKVPPLYFVA 578
Db      939 SKNGSAASKVLLLPDVSALATQAEIESIYKPVLYKAEAEKRLPSDETEKEDRSPSAIFSA 998

RESULT 35
US-09-972-599A-6
Sequence 6, Application US/09972599A
Patent No. US20020077295A1
GENERAL INFORMATION:
APPLICANT: STRITTMATTER, STEPHEN M.
TITLE OF INVENTION: NOCO-RECEPTOR-MEDIATED BLOCKADE OF AXONAL GROWTH
FILE REFERENCE: C077 CIP US
CURRENT APPLICATION NUMBER: US/09/972,599A

```

```

: CURRENT FILING DATE: 2001-10-06
: PRIOR APPLICATION NUMBER: PCT/US01/01041
: PRIOR FILING DATE: 2001-01-12
: PRIOR APPLICATION NUMBER: 09/758,140
: PRIOR FILING DATE: 2001-01-12
: PRIOR APPLICATION NUMBER: 60/236,378
: PRIOR FILING DATE: 2000-09-29
: PRIOR APPLICATION NUMBER: 60/207,366
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/175,707
: PRIOR FILING DATE: 2000-01-12
: NUMBER OF SEQ ID NOS: 57
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 6
: LENGTH: 1192
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-972-599A-6

```

Query Match	2.5%	Score 120.5	DB 10	length 1192
Best Local Similarity	18.2%	Pred. No. 5.2		
Matches 120	Conservative 102	Mismatches 245	Indels 193	Gaps 30

```

QY      2 SKPYLFARSRMPVLAAYVPLMTSQAQAQONPNANIINHVAHDTALNOAKAGNPVLL 61
Db      449 STPEBICKRSGAYITICAFNPAPAT-ESTA-----TNPFLLCDPPTSEKNTDEKKIEE 499
QY      62 TPEOIQARLNAAGLNAPK-----OSQALDVNFDDOSPISRIGEOSPPLGLDMVIEETT 116
Db      500 KKAQIVTEKNTSTKTSNPFVAAQOSETDYTTDNLKFTVEEIVANMBEGLPDLVQENC 559
QY      117 PLSLBEL-----FADES-----TEMGINPNDYIPEYQEQP 147
Db      560 ESELENEVIGTKIAYETKMDLVOTSEVMQESLYPAAQLCPSEESSEATSPVLPJDIYMAP 619
QY      148 -NSEV-----VBPPTLEPEKPGILKILRYARLENDGVAKVPYLRKAKFPQSSOCSETSAIG 200
Db      620 LMSAVPAGASVIGPSSSPLEASSVN--YESIKHPEPNPP-----YEEAMSVSLKAYS 671
QY      201 SSHOKEPEYANIKAALEDITQ--ESAMDNGSIPRLRQATALVARAVGYDI-DLSIIR 256
Db      672 GIKEIKKEPENINAALOETEAFTYSIACDL-----IKETKLSAEPAPDFDSYSEMAKVE 725
QY      257 NSISEVDIIDL---GPPYIYDRAVEYREBGADKAFETY--ADEVPLLIGDVFHG 310
Db      726 QPVDHSELVEDSSPDSPPVLD-----FSDSIPDPVPOKODETMYLVKESLRET 774
QY      311 KYET-----KNLLENASAEHG--YFDG--RWLDRSVVILPDMTADVSLIYDGTQYRF 361
Db      775 SFEEMIEYENKEKLSALPEEGCKPYLEBFKLSLNDTKDLPDEVSITLS----- 823
QY      362 DEVVEFTIDPKTNOLTDPDKLPVKRELLLEOLTIVNNGEAYNLQAVRALNSDLIATRYFN 421
Db      824 -----KKEKIPLO--MEELSTA-----VYSDN----- 843
QY      422 MWNTEIYEPEREQIONOVSEFQSSSRTEPRAQVDESTLEPIETVELTDSILMDISPIE 481
Db      844 -----DLTISKQAQIREME-FTSDSS-----PIETIDEPFLISSKT 879
QY      482 FSAENLIQDKLNLVAAKARHLYDMPDR---VLAINHDDGVNRSILGRISDAV----- 531
Db      880 DSFSEKLARETDLDEVSHKSELIANAPDGAGSLPCHELPHDLSL-KNIPKVEKISFSDDF 938
QY      532 -----SAVARAIL--PD-----ESENVEYIDLP--RTLALNRKTPADVYQSKKVPVLYFVA 578
Db      939 SKNGSATSKVLLPDPVYSALATQAETIESIYKPKVLVKEAEKRLKLPDTEKEDRSPSAIFSA 998

```

APPLICANT: Council of Scientific and Industrial Research
TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
FILE REFERENCE: 063915
CURRENT APPLICATION NUMBER: US/09/820,843A
CURRENT FILING DATE: 2001-03-30
NUMBER OF SEQ ID NOS: 118
SOFTWARE: PatentIn version 3.0
SEQ ID NO 15
LENGTH: 1139
TYPE: PRT
ORGANISM: M. genitalium
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: cytochrome-c-accessory protein
OTHER INFORMATION: 9111046012
US-09-820-843A-15

Query Match 2.5%; Score 119.5; DB 9; Length 1139;
Best Local Similarity 20.3%; Pred. No. 5.7; Indels 235; Gaps 42;
Matches 158; Conservative 113; Mismatches 272;

64 EQIARLNAALNAKPOSOALDVNFDDQSPISRIGESPPGLDMYIEETPLSLEL 123
293 DQVQANDLN--NEEIPTAEKKVTTDESKQAOVVD--SYOLPITDQDDQTFSSSFE- 347
124 FAQSTEMGINPNDIPEYQGEQNPSEVVPPTLEP--EKGLIK-----RLYARLF 173
348 -----TQPTVQFOQVNVSEVNDQFKPEITKEPVSEFNKQDVSTDLNSESNLYSENN 402
174 NDGVNKKVRLAKFYQSSQGETSAIGSSHO--KTEPVANIKAALEDITOESAMDLNGSIP 232
403 KDATNN--DLSLSEFIQLNSNETASDVHYESKSEPHIDYKFG--SDLSQSN--NSISE 457
233 RLROTALVAARAVGYIDLSITRNSIGEVYIHDIGEYVYIDRAVEVREGADDKAF 292
458 -----SEP-----VKFNSETAPDAHF 473
293 TTVADEVPLLCGVFHHCKYETKKNLIENASAEHGYFDGRMLDRS--VDVILPDNTADVS 350
474 ESQSPVQVQVYDIQN--BELKPTLDQPPSSD--DYFAKQPTDEYVGFNDNDLPREYKQPE 529
351 LIYPTGTQYRDEVEVFTIDPK--TNQLTTPDKLP--VKRELEQLLTVMGAEVNLQ 405
530 SVVD--QPSDD--HFAKPESTDSYSPSSD--LPQPTLDQPSLDHNVQVYFHDHELK 583
406 AVRALS-----DLIATRYNMVTEIVFPEREQ-----IQ 436
584 PVAEQNNYQVGFQVOA--NLDNNEIOTPAKEVTTDESKQAOVVDSYOLPITDQ 640
437 NDQVSEBSSSRTEPA-----QVD-----ESTLEPVLET-----VELTGLIMD 476
641 QDQTTT--SSSFEIOTPEQFDQVSEVNDQFKPEITKEPVLESFNKQDVETENTYNN 698
477 ISPIEFSANLI-----QDKLNLVAARAHLYDMPDDRVLAINHDDGVNRSI--LGR 526
699 LQKFDIGDNKITITTKSSPOIPTPLPISFVSNRIEYKPVETLALDNKESQEQITINS 758
527 ISDAVASARALPDSENEVYIDLPERTALNRK--TPADYQSK----- 570
759 ITESKTLAKTL-----SVQLQOINSLNQSIVTSESVALDKKDDQLTINTVNSEQ 810
571 -VPLVYFVASDKPRDQIGLQGWSDGTRLVTKFEHNLII--NRDGYQAGAELR----- 620
811 QPKIEVFPAKAEPE-----EHSITQNKQSVEDSELDNFKKSDL 851
621 ---ISEDKKGVKLYATKPLSHPLN--DQLRATLGYQOEY--FGHSTNGF-----DLSTR 667
852 YKIISLEKRG-----ELNPTINFDAIFQMWNDYQMSVQSFIL--LNFVTVYKKNQISER 903
668 -----TLEHISRIITNGCMNRTYSLRYRLDKLKTQAPPTWDDLPVDF--VKNKPS 718

Db 904 YLIKKELQSELSRLIDQENENLVQFNNAKNLTTLOKE---EMIRSLASDAFAIVKPS 958

RESULT 37
US-09-815-242-5480

Sequence 5480, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlisen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes In

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows version 4.0

SEQ ID NO 5480

LENGTH: 1160

TYPE: PRT

ORGANISM: Staphylococcus aureus

US-09-815-242-5480

Query Match 2.5%; Score 119.5; DB 10; Length 1160;
Best Local Similarity 18.8%; Pred. No. 5.8;
Matches 169; Conservative 116; Mismatches 279; Indels 333; Gaps 42;

108 DMSVTEETPLSLELFAQSTEMGINPNDYI-----PEYQGE-----QPN 148

346 DETLEIEGPVYFSLFA--TTMPKLNHIITKFSCKPVQGFQGYDINRSEFORVYNN 402

149 SEVVVPPPLEPPEKPGIKLYARLFNDGVNKKVRLAKFYQSSQGETSAIGSSHQTEP 208

403 YIIIVLVETEIK-----VERMQAMLESMHPSI-----TKIHRMSSQQAIIIEGS--LSGG 452

209 VANIKAALEDITOESAMDLINGSIPRLQTLVAARAV--GYVIDLSTIRNSIGEVYITH 267

453 FELPDMGLVITVERELFKSKQKKRRTKAISNAEKIKSYODLNVG-----DYIWH 503

268 DLGEFVYIDRAVEVREGADDKAFITVADEVPLLCGVFHHGKYEETKKNLIENASAEH 327

504 -----VHHGV----- 508

328 YPDGKMLDRSVVILIPDNTADVSLIYDGTQYRFDEVVEFTID--PKTNQLTTPDPKLPV 385

509 ---GRVIGVETIEVQGTNRDIYKL-----QYKGTQDLFVPDQDMQVQKYVASSEDKTPK 559

386 KREL--LEQLLTVMNGEAVYNLQAVRALSNDLIATRYENNVNTEIYFPEREQIONDQVSTE 443

560 LNKLGSEWKKT-----KAKVQOQSVEDIAELIID-----LYKEREM----- 595

444 QSSSSKTEPAQVDESTLEVIETVELTDSILMDISPIESASNLIDQKLNLAARAHLY 503

[illegible]

```
APPLICANT: PRINJHA, RABINDER KUMAR
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GP-30165-C1
CURRENT APPLICATION NUMBER: US/09/789,386
CURRENT FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: U.K. 9916898.1
PRIOR FILING DATE: 1999-07-19
PRIOR APPLICATION NUMBER: U.K. 9816024.5
PRIOR FILING DATE: 1998-07-22
PRIOR APPLICATION NUMBER: US 09/359,208
PRIOR FILING DATE: 1999-07-22
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1192
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-789-386-2
```

```
Query Match 2.5%; Score 119.5; DB 10; Length 1192;
Best Local Similarity 18.0%; Pred. No. 6.1;
Matches 116; Conservative 99; Mismatches 241; Indels 187; Gaps 29;
```

```
QY 19 AYLPMTSQALAOONNPANIINHVAHDTAIQAAGNPVLLTPQIQAARLNAAGLNAAK 78
DB 460 AYTCAPFNPATTESTATINIF---PLGDPITSEKNTDEKKIEKKAQIYTEKNTSTKTSN 516
QY 79 P-----QSQALDVNFDOSPIRIGESQSPGLDMSVIEETTPLSLEEL----- 123
DB 517 PELVAQOSETDYVTVDNLTKYTEEVANMPGLTPDLVOACESELNEVTGKIAVETK 576
QY 124 -----FAQES-----TEMGINPDNDIPEYGEOP-NSEY-----VVPPTL 157
DB 577 MDLVOTSEVMQESLYPAQOLCPSEFESEATPSPVLPDIYMEAPLNSAVPSAGASYIOPSS 636
QY 158 EPEKGLIKRLYARLFNDGVNKPRLKAKFYOSSQSGETSAIGSSHQKTEPYANIKAALE 217
DB 637 SPLKSSVN--YESIKHEPNRP-----YEAMSVLKKYSGIKELKEPENINAAQ 688
QY 218 DITQ---ESAMDLNLSIRLROTALVAARAVGYDI-DLSITRNSIGEVYIIHDL---G 270
DB 689 ETEAPYISICDL-----IKETKLSAEPAPDFSDYSEMAKVEQVPDHSSEIVEDSSPDS 742
QY 271 EPPVYIDYRAVEVRGEGADDKAFTV---ADEVPLDIGVFHNGKYEY-----KKNLJENA 322
DB 743 EPPVLDL-----FSDDSIPDVPOKODEYVMLVKESLTETSESMIEYKKEKLSAL 791
QY 323 SAHNG--YFDG--RWLDRSVYDILPDNTADVSLIYDTGYRFEDEVVFTTIDPKTNQLT 378
DB 792 PPEGKPYLESFKLSLDNTKDTLLPDEVSTLS----- 823
QY 379 DDDKLPVKKELLEQLLYVMGAEYVNLQAVRALSNDLIATRYENMVTETVFPFERQIQND 438
DB 824 KKEKIPLO---MEELSTA-----VYSND-----DLFISKEAQIRRT 856
QY 439 QVSFEQSSSRTEPAQOVDESTLEPIETVELTDGILMDISPIEFSASNLIDPKLNLVAAK 498
DB 857 E-TFSDSS-----PIETIDEPFTLISKTDSFKLAREYTDLEVSH 896
QY 499 ARHLVMPDDR---VLAINHDDGVNRSILGRISDAV-----SAVARAIL--PD- 541
DB 897 KSEIANAPDAGSLPCTELPHDLSTL-KNIQPKVEEKISFSDFSKNGSATSKVLLLPDV 955
QY 542 -----ESENEVIDLPE-RTALANKRTPADYQSKVPLVYFA 578
DB 956 SALATQAEIESIVKPKVLVKEAEKKLPDTEKERDPSAIFSA 998
```

```
RESULT 40
US-09-893-348-23
Sequence 23, Application US/09893348
Patent No. US20020072493A1
GENERAL INFORMATION:
```

```
APPLICANT: EISENBACH-SCHWARTZ, Michael
APPLICANT: COHEN, Irun R.
APPLICANT: BESERMAN, Pierre
APPLICANT: MOSONOGO, Alon
APPLICANT: MOALEM, Gila
TITLE OF INVENTION: ACTIVATED T-CELLS, NERVOUS SYSTEM-SPECIFIC ANTIGENS AND THEIR
FILE REFERENCE: EIS-SCHWARTZ-2A
CURRENT APPLICATION NUMBER: US/09/893,348
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 09/314,161
PRIOR FILING DATE: 1999-05-19
PRIOR APPLICATION NUMBER: US 09/218,277
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: PCT/US98/14715
PRIOR FILING DATE: 1998-07-21
PRIOR APPLICATION NUMBER: IL 124500
PRIOR FILING DATE: 1998-05-19
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn version 3.1
SEQ ID NO 23
LENGTH: 1192
TYPE: PRT
ORGANISM: Homo sapiens
US-09-893-348-23
```

```
Query Match 2.5%; Score 119.5; DB 10; Length 1192;
Best Local Similarity 18.0%; Pred. No. 6.1;
Matches 116; Conservative 99; Mismatches 241; Indels 187; Gaps 29;
```

```
QY 19 AYLPMTSQALAOONNPANIINHVAHDTAIQAAGNPVLLTPQIQAARLNAAGLNAAK 78
DB 460 AYTCAPFNPATTESTATINIF---PLGDPITSEKNTDEKKIEKKAQIYTEKNTSTKTSN 516
QY 79 P-----QSQALDVNFDOSPIRIGESQSPGLDMSVIEETTPLSLEEL----- 123
DB 517 PELVAQOSETDYVTVDNLTKYTEEVANMPGLTPDLVOACESELNEVTGKIAVETK 576
QY 124 -----FAQES-----TEMGINPDNDIPEYGEOP-NSEY-----VVPPTL 157
DB 577 MDLVOTSEVMQESLYPAQOLCPSEFESEATPSPVLPDIYMEAPLNSAVPSAGASYIOPSS 636
QY 158 EPEKGLIKRLYARLFNDGVNKPRLKAKFYOSSQSGETSAIGSSHQKTEPYANIKAALE 217
DB 637 SPLKSSVN--YESIKHEPNRP-----YEAMSVLKKYSGIKELKEPENINAAQ 688
QY 218 DITQ---ESAMDLNLSIRLROTALVAARAVGYDI-DLSITRNSIGEVYIIHDL---G 270
DB 689 ETEAPYISICDL-----IKETKLSAEPAPDFSDYSEMAKVEQVPDHSSEIVEDSSPDS 742
QY 271 EPPVYIDYRAVEVRGEGADDKAFTV---ADEVPLDIGVFHNGKYEY-----KKNLJENA 322
DB 743 EPPVLDL-----FSDDSIPDVPOKODEYVMLVKESLTETSESMIEYKKEKLSAL 791
QY 323 SAHNG--YFDG--RWLDRSVYDILPDNTADVSLIYDTGYRFEDEVVFTTIDPKTNQLT 378
DB 792 PPEGKPYLESFKLSLDNTKDTLLPDEVSTLS----- 823
QY 379 DDDKLPVKKELLEQLLYVMGAEYVNLQAVRALSNDLIATRYENMVTETVFPFERQIQND 438
DB 824 KKEKIPLO---MEELSTA-----VYSND-----DLFISKEAQIRRT 856
QY 439 QVSFEQSSSRTEPAQOVDESTLEPIETVELTDGILMDISPIEFSASNLIDPKLNLVAAK 498
DB 857 E-TFSDSS-----PIETIDEPFTLISKTDSFKLAREYTDLEVSH 896
QY 499 ARHLVMPDDR---VLAINHDDGVNRSILGRISDAV-----SAVARAIL--PD- 541
DB 897 KSEIANAPDAGSLPCTELPHDLSTL-KNIQPKVEEKISFSDFSKNGSATSKVLLLPDV 955
QY 542 -----ESENEVIDLPE-RTALANKRTPADYQSKVPLVYFA 578
DB 956 SALATQAEIESIVKPKVLVKEAEKKLPDTEKERDPSAIFSA 998
```

Tue Apr 29 12:22:31 2003

us-09-914-168-2.rapb

Page 29

Search completed: April 28, 2003, 16:32:38
Job time : 97 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:19:19 ; Search time 51 Seconds
(without alignments)
1732.305 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 4727
Sequence: 1 MSKPYLFANRSEMPVALAAY.....TGVEEGNPDKLHFFIGRPF 919

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pIR_73:*\br/>2: pIR1:*\br/>3: pIR2:*\br/>4: pIR3:*\br/>5: pIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	615.5	13.0	579	2 F83327	conserved hypotnet
2	494	10.5	578	2 AE0428	probable exported
3	479	10.1	577	2 AD1055	probable exported
4	454	9.6	582	2 F82064	conserved hypotnet
5	436.5	9.2	578	2 B64012	hypothetical prote
6	433.5	9.2	577	2 S56445	hypothetical 64.8K
7	433.5	9.2	577	2 F91278	hypothetical prote
8	433.5	9.2	577	2 F86119	hypothetical prote
9	347	7.3	617	2 G87206	conserved hypotnet
10	259	5.5	797	2 D82000	outer membrane pro
11	259	5.5	797	2 G81228	outer membrane pro
12	236.5	5.0	617	2 AE2897	conserved hypotnet
13	236.5	5.0	641	2 H97672	hypothetical prote
14	232.5	4.9	774	2 D97527	hypothetical prote
15	232.5	4.9	774	2 AE2746	group 1 outer memb
16	225	4.8	861	2 S77409	outer membrane pro
17	223.5	4.7	781	2 AH3355	outer membrane pro
18	220	4.6	739	2 A82025	probable outer mem
19	216.5	4.6	628	2 A81430	outer membrane pro
20	216	4.6	628	2 B87448	outer membrane pro
21	215	4.5	635	2 G81003	conserved hypotnet
22	214	4.5	784	2 E82731	outer membrane ant
23	211.5	4.5	795	2 AC0129	conserved hypotnet
24	209	4.4	623	2 A13488	outer membrane su
25	206.5	4.4	916	2 G64601	outer membrane pro
26	205.5	4.3	768	2 B97725	outer membrane pro
27	202.5	4.3	768	2 B97725	outer membrane pro
28	197.5	4.2	833	2 AE2089	hypothetical prote
29	194.5	4.1	906	2 F71910	probable outer mem

30	193.5	4.1	676	2 AE2417	hypothetical prote
31	190.5	4.0	769	2 F87486	outer membrane pro
32	183.5	3.9	797	2 H83190	probable outer mem
33	178	3.8	810	2 A64742	hypothetical prote
34	178	3.8	810	2 C90651	hypothetical prote
35	178	3.8	810	2 C85502	hypothetical prote
36	177.5	3.8	778	2 C70412	outer membrane pro
37	177.5	3.8	803	2 B82099	surface antigen VC
38	171.5	3.6	797	2 JC4078	protective surface
39	171	3.6	853	2 D70304	hypothetical prote
40	170.5	3.6	588	2 AD2445	hypothetical prote
41	168	3.6	808	2 F64102	protective surface
42	159.5	3.4	1319	1 C43735	bcsC protein - Ace
43	159	3.4	475	2 T11586	hypothetical prote
44	152.5	3.2	491	2 AC1816	hypothetical prote
45	149.5	3.2	803	2 AB0530	outer membrane pro

ALIGNMENTS

```
RESULT 1
F83327
conserved hypothetical protein PA2543 [imported] - Pseudomonas aeruginosa (strain PAO
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83327
R:Stover, C.K.; Pham, X.Q.; Ewlin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardy, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; M0ID:20437357; PMID:10984043
A:Accession: F83327
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-579 <STO>
A:Cross-references: GB:AE004682; GB:AE004091; MID:99948598; PIDN:AA05931.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2543

Query Match          13.0%  Score 615.5; DB 2; Length 579;
Best Local Similarity 24.5%; Pred. No. 1.8e-27;
Matches 178; Conservative 117; Mismatches 220; Indels 213; Gaps 17;

QY 210 ANIKALEDTOPSAMDLSIPRLROTALV---AARAVGYDIDLSIRNSIGEDVLI 265
      |||:| : : : : |||:| : : : : |||:| : : : : |||:| : : : :
DB 41 ANIEAVVSLGERD---EALQRFRRNMEAQAQKAAQALGTYQ-----AQIDSE 86

QY 266 IHD-----LGPVYIDYRAVEVNGEGADDKAFTTVADENVPLIGDVFHNGKET 314
      :| : : : : |||:| : : : : |||:| : : : : |||:| : : : : |||:|
DB 87 VKDGKPKKTLTKVPCPEVRLRQVINQVIGEAASLSEFRLPSK-QLKPGAKLNQGVYED 145

QY 315 KKKLILNASEHGYPDGRMLDRSVNVLITDNTADVSLIDTQGYRDEVFETTPDKTN 374
      |||:| : : : : |||:| : : : : |||:| : : : : |||:| : : : : |||:|
DB 146 AKRLIONQASRYGFQGRSTORLSIDPRAGIADIDVDSGGRYTFGKVSF----- 197

QY 375 QLTTPDKLPVKKRELLQGLTVNMGAVNLQAVRALSNLILARYNMMNTLIVPEREQ 434
      | : : : |||:| : : : : | : : : |||:| : : : : |||:| : : : : |||:|
DB 198 ----DQDSI-IEEELLRRVVPFRAGOPYDSELIAELNQNQSSGYF----- 238

QY 435 IONDVSEFOSSSSRTPEAQVDESTLEPIETVELDGLIMDISITIEFSANLIDDKML 494
      : : : : |||:| : : : : |||:| : : : : |||:| : : : : |||:|
DB 239 -----EGVRVDAAPYQAQ----- 252

QY 495 VAAKAKHLVDMDDRLALNHDGVNRSLIGRISDAVSAVARAILPDESENEYIDLPERT 554
      : : : : |||:| : : : : |||:| : : : : |||:| : : : : |||:|
DB 253 -----DGAQAI----- 259

QY 555 ALANRTPADVYQSKKVPPLVVFVASDPKPDGQIGICGWSGDTGRLTATKEHNLINRDGYO 614
      : : : : |||:| : : : : |||:| : : : : |||:| : : : : |||:|
DB 260 -----PVAVRLKARKPTMTGVLGFTSDVGAARAFNWTIRHVNVEGHS 302
```

QY	615	AGAEIARLSEDDKKYKLVATKPLSHLPNDOLATATGQOEVRGSHNPNGLDLSRRT-----E	670
Db	303	LGFESEIAPRPNONGAMETIEPLDDPLTDLKRTSGTOPE-----DLVDTESKLLTLGGE	356
QY	671	HEISRSIIIONGGWNRVTSLRYRLDKIKTQAPPETMODLPLVDFVNGKPS--OEALLAGVAV	728
Db	357	WHSKR-----GGMGRVVSFLNMWREYK-----GDDSGLSLFLMPGIIcy	396
QY	729	HKTVADNLVNMRCGRORYSLEVGSSGLVSDANMALIARAGISGVYSEFGDNAYGNSRAHOM	788
Db	397	SLEETDNKVDVDSHGTRLOPNNYKGAKEGLLADADV.LHYDAMAGLTSFAG-----GHL	449
QY	789	TGGIOAGYIWSDNENHVPYRLREFAGSDOSIRGVAHSDLSPTSDKGYTLGGVAVLAVTAE	848
Db	450	LGRLÖVGGIATNDYKSIPIPSLRFAGGDQSVRGYDYLTLSPENSQGDKICGGRYTAGSYVE	509
QY	849	YNPEEMKCLRLAVFGDIDINADKGFNTDKIGAGVGRMASVGVGVAVDVARGVKEEGNP	908
Db	510	YQYPLAEWMRLATPFDOGNATFNSLDFPSIKTYGVGVRKWSVYVGLRDLDAIALADDDG-	568
QY	909	IKLHFFIG 916	
Db	569	FLRHESMG 576	

[illegible]

```

OY 565 VYQSKVPLVYFVASDKPRDGOIGLGMGSDPTGTRLYTFEEHNLINRPOYQAGAEI,SED 624
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 -----ELGGYATVDGPRILTASMKRPMWNSGHSILTTTALTASAP 308
OY 625 KKGVL-YATKPLSHPLNDLRLATLGTQOEYFVGHSHNGCEDLSTTLEHLSRSLIQNGW 683
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 309 EQLTFESYRILRNPRLNPEOYLLIOGFRR-----TDLNPNSTPTLLNARFWDLSGGW 362
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 684 NRTYSLRRLRLKLTQAPPEPTMODLPVFNKGKESQELT--ACVAAYKHTVADMLVPMR 741
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 363 QRAILRRSLD-----HFQGRYDTTMLLYPCVSLNRTQROGAMPVW 406
OY 742 GYRQYSLLEVSSGLVSDANNAIARAGISGYISFGDNAYGSNRHQMOTGGTQAOYIMSDN 801
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 407 GDSQRYSDIVSDPTWGSVDGIFDQONWMTRTLGEK-----NRFVARGNWVIEFINN 459
OY 802 FNHHYRLRFEPAGDQSIIRGAHDSLPSIDSKGYLTGOVLAITYAEVNEFFMMDRLYAR 861
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 460 FDRVPSLRFLFAGDGRSIRGYKFNRIISPDRDCKLTGASKLATSLEYQYFNGRMWCAV 519
OY 862 FGDIGNAYDKGFTNDTKIGAGYGVMAASPVQGVARYDAVTGKEGNSPKLHFFIG 916
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 520 FVDSEAVANNRSKSLDKTGAGYGVMAASPVKPKIDIAAPL-GDNEHTHGVQFYIG 573

```

```

RESULT 3
AD1055
problemable exported protein ytfM [imported] - Salmonella enterica subsp. enterica serov
C:Species: Salmonella enterica subsp. enterica serovar Typhimurium
A>Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AD1055
R.Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T., Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
, S.; Moule, S.; O'Gaora, P.
Nature 413, 846-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: AD1055
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-577 <PAR>
A:Cross-references: GB:ALU51382; PIDN:CAD06889.1; PID:g16505537; GSPDB:GN00176
C:Genetics:
A:Gene: ytfM

Query Match          10.1%; Score 479; DB 2; Length 577;
Best Local Similarity 22.7%; Pred No. 1,1e-19;
Matches 166; Conservative 104; Mismatches 257; Indels 204; Gaps 17;

QY   210 ANKAALEDITQESAMDNGSIPLRLQTALVAAR-----AVGYD-IDL 252
      |||::||::||::||::||::||::||::||::||::||::||::||
Db    22 ANVRUKVEGLSGELKEKNRAQLSTIQSDVMPDRFRFARVDATREGLKALGYEPTIKF 81

QY   253 SIIRNSIEVEYIITHDL--GEPVYIDRAYEVREKGADKKFTYADVDPVLLIDVYHHG 310
      ::||::||::||::||::||::||::||::||::||::||::||
Db    82 DLLPPAKRGROVLIAVRYPGGPVLIGTGVTLRGARTDKYLALLTRP-AIGTVLNQG 140

QY   311 KYETPKNLIIENASAEHGFEDEGRMLDRSYDVLLIPNTADVSILPYTGQYRFDEVVFETID 370
      |::||::||::||::||::||::||::||::||::||::||::||
Db    141 DYDNKRKSGLTSVLKGYDFDSFIKSQLGIAMLRQAQAWMDIDYDGSERGRRGPVTF---- 196

QY   371 PKTNOLTTDDPKLPVKRRELLEQLLTVNNGEAYNLCAVAVALSNDLIATRYFMNVMTIELVP 430
      :::||::||::||::||::||::||::||::||::||::||::||
Db    197 -----ESSQIRDEYLONLPRFKEGDYESKDIALNRRLRSATGWFSV---VVAP 243

QY   431 EREQIQNQVQSFEQSSSRTEPAQVDESTLEPIETVELTDGILIMDISPIEFSSANLIQD 490
      |::||::||::||::||::||::||::||::||::||::||::||
Db    244 E-----PEKSKRKTIIP----- 255

QY   491 KLNIIVAAAKARHLYDMPPDDRVLAINHDDGVNRSILGRISDAVASARAAILPDSENEVIDL 550

```

Db 256 -----LKGVSPTNTI--- 268
OY 551 PERTALANRKPADVQSKKVPFLYVAVASDKPRDQIGLGWSDPTGRLVTKFEHLINR 610
Db 269 -----ETGVGSTDVGRVAVASMKKPPMNS 293
OY 611 DGQAQAEELRLSEDKKGVKL-VATKPLSHPLNDLRLATLGYQOEVGHSTNGFDSLTRL 669
Db 294 YGHSJLTSTISAPAEQYLDPSYKPKLLKNPLEQYLLVQGGFKRT---DLNDTQDQSTTL 349
OY 670 EHLISIIIONGGMNRRYSRLRYLDKLTQAPPETMODLPEVFNKPSOEALL--AGVA 727
Db 350 --AVSRWDLSSGMORINLRMSFD-----HFQGVNTTMTMLFYGVGM 391
OY 728 VHKTVADNLVPMRGYRQRYSLSEVSSGLVSDAMALARAAGISGVSEGDNAVGSNRAHQ 787
Db 392 ISSTRBGRGLMPWGDQSRISVDYSNTAMGSDVDVSVLQAKONWIRFLYDR-----HR 444
OY 788 MTGGIAGYIWSDFNHNVPYRLRFPAGGDSIRGYAHDLSPLSDKGYLTGGQVLAAGTA 847
Db 445 FVVRANLGIETGDFDKVPPDLRFPAGGDSIRGYKXKISPKXSDGNLKGASKLATGSL 504
OY 848 EYNYEFMKDLRLAVFGDIGAYDKGFTNDTKIGAGVGRMASPVGOVAVD--VATGVKEE 905
Db 505 EYQYNTGKMGWAVFVDSGEAVSDIRSRDFKTGTGVGRMASPVGKLDFAVAVGDKDE 564
OY 906 GNPRLKHFIFIG 916
Db 565 HG---LQFYIG 572

RESULT 4

F82064
conserved hypothetical protein VC2548 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82064
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoli, I.; Sellers, J.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Mature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; M0ID:20406833; PMID:10952301
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-582 <HEIDI>
A:Cross-references: GB:AE004323; GB:AE003852; NID:9657119; PIDN:AAF95689.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2548
A:Map position: 1

Query Match 9.6%; Score 454; DB 2; Length 582;
Best Local Similarity 22.4%; Pred. No. 3e-18;
Matches 164; Conservative 104; Mismatches 237; Indels 228; Gaps 20;

OY 211 NIKAALEDITQESAMDLNLSI---PRLROTALVAARAVGY--DIDLSTRNS----- 258
Db 48 NVEAYLSTI--AAODYSTSLRFQSLERSMTEALNALGYHPSIFTVSEDNQRLRAAV 104
OY 259 -----IGEVVDYIHDLDEPVYIDYRAVEVREGADKAFTTVADEVPLLIDVFNHGK 311
Db 105 TLGEVYTRLESEVDIYI-----KEAGSDRDFORLIRRSGRVAPALNHS 148
OY 312 YETKKNLLENASAEHGYFDGRWLDRLRSVDYILPONTADVSLTYDTGYRDEVEVFFTIDP 371
Db 149 YDNLSKGINLNLQKGYFNGDFQASRLVLEIPELNQARVILIHPSGIRLYL----- 198
OY 372 KTNQULTDPPDKLPVKRELLLEQLLTVNMGAEAYNLQAVRALSNDLIATRYENMVTLEIPE 431
Db 199 --GATTVESQIDENKVM--SLRPFKQGEPIYLVQGEFNQNLNSNTDWFSSVVE---PD 251

OY 432 REQIIONDVSEQSSSSSRTEPAQVDESTLEBPVIEVTELTGILMDISPIERSASNLQDK 491
Db 252 LSQI----- 255
OY 492 LNLVAARHLIYMPDDBRVLAINHDDGVNRSILGRISDAVAVARAILPDESENEVIDLP 551
Db 256 -----DEG----- 258
OY 552 ERTALANRKPADVQSKKVPFLYVAVASDKPRDQIGLGWSDPTGRLVTKFEHLINR 611
Db 259 -----RELPIKVTALAPQARNOLETGIGYSTDVGVRSGLMVKGPVNSQ 301
OY 612 GYQAQAEELRLSEDKKGVKLVAATKPLSHPLNDLRLATLGYQOEVGHSTNGFDSLTRL 670
Db 302 GHSDDSSPSLSIPQITTAGIKIPLDALNRYIYQYGMK-----LDKRDTESE 352
OY 671 HEIS--RSIIIONGGMNRRYSRLRYLDKLTQAPPETMODLPEVFNKPSOEALLAGVAV 728
Db 353 SNLSLERHMDLGGMHRTVFIRYLLENYR-----QGLQDD-----NSQFLPGMTY 398
OY 729 HKT--VADNLVPMRGYRQRYSLSEVSSGLVSDAMALARAAGISGVSEGDNAVGSNRAHQ 787
Db 399 TRTTRNSGGLTWGDKQYITLLEYGDPALLSETRVLRLOTGSSWMLRYARN-----HR 451
OY 788 MTGGIAGYIWSDFNHNVPYRLRFPAGGDSIRGYAHDLSPLSDKGYLTGGQVLAAGTA 847
Db 452 ALVAVDGGANLVDFEQSLRFPAGGDNLRGYKXKISPDASGALTAKYATVSSI 511
OY 848 EYNYEFMKDLRLAVFGDIGAYDKGFTNDT--RTGAGVGRMASPVGOVAVDVAATGV-K 903
Db 512 EYQYRLTGNNMAAFMVDGAF-----NDNEPKKGYGTGIRWISPGIRLDFRWAGLDA 566
OY 904 EECNPRLKHFIFIG 916
Db 567 APGDEFKIHFTIG 579

RESULT 5

B64012
hypothetical protein H10698 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: B64012
R:Reischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; M0ID:95350630; PMID:7542800
A:Accession: B64012
A:Status: nucleic acid
A:Molecule type: DNA
A:Residues: 1-578 <TRIGR>
A:Cross-references: GB:U32752; GB:U42023; NID:91573692; PIDN:AAC22357.1; PID:91573700

Query Match 9.2%; Score 436.5; DB 2; Length 578;
Best Local Similarity 20.5%; Pred. No. 3e-17;
Matches 142; Conservative 105; Mismatches 235; Indels 211; Gaps 14;

OY 243 RANGYDIDLSIIRNS-IGEVVDYIHDL--GEPIYIDYRAVEVREGADKAFTTVADEV 299
Db 73 RVEGYESSVRFERKORGRDLIAHVTEGEPKKTAGTVOJEGEAAODENFALKNL 132
OY 300 PLLIGDVFNHGKYEYTKKNLLENASAEHGYFDGRWLDRLRSVDYILPONTADVSLTYDTGY 359
Db 133 P-KDGYLVEHQYTDYKTAISRLALNRGTYFDGNFKISRLSEPTHQAMRMRLDSSVRY 191
OY 360 RDEVEVFFTIDPPTNDLTTDPDKLPVKRELLLEQLLTVNMGAEAYNLQAVRALSNDLIATRY 419
Db 192 HYNITF-----SHSQIRDDYLVNINLIKSSDDPYLMNLSLDTSDPFSNM 237

```
QY 420 FNNVNTTIVPEREIOIONDOVSFEQSSSRTEPAQVDESTLEPIETVELTDGILMDSP 479
|:|
Db 238 FSSV----- 241
QY 480 IEFASNLIDQKLNLAAKARHLVMDPDDRLAIINHDDGVNRSILGRISDAVASARAIL 539
|:|
Db 242 -----LVQPNV----- 248
QY 540 PDESENEVIDLPERTALANRKTTPADYQSKKVPILYEVASDKPRDQOIGLQWSDTGTRL 599
|:|
Db 249 -----HKSKTVDYEIILLPRKKNAMELVGFSTDGCVHG 282
QY 600 VTKREHNLINRDYQAGAEIRLSEDKKGVK-LYATKPLSHPLNDQLRATLGQOEVEGHS 658
|:|
Db 283 QIGTKFWINSRGLSLRSNLTLSPKOTLEATFKMPLKLNLYDFAVWME---GEK 338
QY 659 TNGFDLSTRLLEHISRSIIIONGMNRTYSLRYRLDKLTOAPPEPTMODLVEFVNGKPS 718
|:|
Db 339 EN-DTMTRVLTLSALRYWNNAHQWQYFGGLRMRYDSF-ITQAD----- 378
QY 719 QEALLAGVAVHKTYADNLVNPMDGYR-----QRYSLVEGSSGLVSDANMA 763
|:|
Db 379 -----ITDKTL---LYPTVGFTRRLRGSEFATWQYOKITFDSLKRIMWSESEFI 427
QY 764 IARAGISGVYSGDNMAYGSNRAHQMTGGIAGYIMSDNFNHPYRLRFAGDGDSIRGYA 823
|:|
Db 428 KVOASAMVRYTAEN-----HRVVARAEIGYLTHTKIEKIPPLTRFFAGDSDSVRYTG 480
QY 824 HDLSLPSIDSKYLTGOVLAVGTAEVNTEFMKDLRLAVFGDIGNAYDKGTNDTKIGAGV 883
|:|
Db 481 YKKIAPKNRNCKLTVGGSRLLTTSLEYQYQVYPMNMAATFFADSGLAADNYATKELRYGTGV 540
QY 884 GVRWASPVGQVRVDYATGVKEGNEPKILHFIIG 916
|:|
Db 541 GVRWASPVGALKEIDIAPIPRDKNSKNIQFYIG 573
```

RESULT 6

S56445

hypothetical 64.8K protein (msra-cbpl intergenic region) - Escherichia coli (strain K-12)

N:Alternate names: hypothetical protein 0577

C:Species: Escherichia coli

C>Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 01-Mar-2002

C:Accession: S56445; G65233

R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.

Nucleic Acids Res. 23: 2105-2119, 1995

A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.

A:Reference number: S56314; MUID:9534362; PMID:7610040

A:Accession: S56445

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-577 <NR>

A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97116.1; PID:g537061

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.: Rose, D.J.; Mau, B.; Shao, Y.

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G65233

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-577 <BLAT>

A:Cross-references: GB:AE000493; GB:U00096; NID:g2367360; PIDN:AAC7177.1; PID:g1790666;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: yclM

A:Start codon: GTG

Query Match

Best Local Similarity 22.9%; Pred. No. 4.4e-17; Matches 164; Conservative 100; Mismatches 262; Indels 189; Gaps 18;

```
QY 211 NIKAALEDITQESAMDLNGSIPRLQRTALVAAVAGY-----DIDLSTIIRNSIGVDYII 266
|:|
Db 38 NVRAQLSTESDEVTPPRRFRARVDDAIRREGKALAGYQPTIEFLRPPPKRGQVLLAK 97
QY 267 HDLGEPPVITYRAVEVRREGADKAFTTVADEVPLLIIDVFHHGKETKLNLENSAEH 326
|:|
Db 98 VTPGVPLVIGGTDVLRGGARDKDYALKLDTFR-AIGVLNQGQYENPKSLSTALRK 156
QY 327 GYFDGRWLDRSYDYLTPNTADVSLIYDTGTOYRDEDEVVEFFIIDTKTQDLTTPDKLKYK 386
|:|
Db 157 GYFDESEFTKADLGALGLHKAFWMDIDYNSGERYRGHVTF-----EGSOIR 202
QY 367 RELLEQLLTVMNGEAYNLQAVRALSNDLIATRFYFMVNTVEIYPEREIOIONDOVSFEQSS 446
|:|
Db 203 DEYIOLNLYPEREGDEYESKDIAELNRLSATGWFNS---YVAPQ-----FDKAR 249
QY 447 SSRTEPAQVDESTLEPIETVELTDGILMDISPIEFASNLIDQKLNLAAKARHLVDM 506
|:|
Db 250 ETKVLP-----LT-GV---VSP----- 262
QY 507 DDRVLAINHDDGVNRSILGRISDAVASARAILPDESENEVIDLPERTALANRKTTPADY 566
|:|
Db 263 -----RTENTTI----- 268
QY 567 QSKKPLYEVASDKPRDQOIGLQWSDTGTRLYTKFEHNLINRDYQAGAEIRLSEDKK 626
|:|
Db 269 -----ETGVGSTDGVPRVYKATWKKPMNNSGSHLTSTISAPQ 309
QY 627 GVKL-YATKPLSHPLNDQLRATLGQOEYFGHSTNGFPLSTTLEHISRSIIIONGMNR 685
|:|
Db 310 TLDPSYKMKPLLNPLEQYLYVOGGFKRT---DLNDTSDSTTL---VASRYMDSLSSQOR 363
QY 686 TYSLRRLRDKLKTQAPPEPTWDLPYDFVNGKPSQALN-AGVAAHKTYADNLVNPMDGY 743
|:|
Db 364 AINLRWSLD-----HFTQGEITNTMLPYFCVMISKRISRGGLPTMGD 407
QY 744 RORYSLEVGSSGLVSDANMAIARAGISGVYSGDNMAYGSNRAHQMTGGIAGYIMSDNFN 803
|:|
Db 408 SQRSIDYSNTAMGSDVDFSVFQAQNNWIMRTLYDR-----HRVTRGTGLGMIETGPD 460
QY 804 HVPYRLRFAGGDSIRGYAHDLSLPSIDSKYLTGOVLAVGTAEVNTEFMKDLRLAVFG 863
|:|
Db 461 KVPDPLRFAGGDSIRGYAKYSIAPKANGDLGASKRLITGSLSYQYVNTGKMWGAFFV 520
QY 864 DIGNAVYDGTNDTIGAGVGRWASPVGQVRVDYATGV-KEENNPILKLFHFIIG 916
|:|
Db 521 DSGEAVSDIRSDRFTGTGTVGVRWESPVGPIKLDRAVPYADKDHG---LQFYIG 572
```

RESULT 7

F91278

hypothetical protein ECs5198 [imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: F91278

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8: 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: F91278

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-577 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA938621.1; PID:g13364675; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs5198

Query Match

Best Local Similarity 22.9%; Pred. No. 4.4e-17; Matches 164; Conservative 100; Mismatches 262; Indels 189; Gaps 18;

[illegible]

RESULT 8
F86119
hypothetical protein ytfm [Imported] - Escherichia coli (strain O157:H7, substrain EDL93)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F86119
R:Perena, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimatala, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: F86119
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1:577 <STO>
A:Cross-references: GB:AE005174; NID:g12519219; PIDN:AAG59418.1; GSPDB:GN00145; UWGP:Z58
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
C:Gene: ytfm

Query Match	9.2%;	Score 433.5;	DB 2;	Length 577;
Best Local Similarity	22.9%;	Pred. No. 4.4e-17;		
Matches 164;	Conservative 100;	Mismatches 262;	Indels 189;	Gaps 18

[illegible]

RESULT 9
G82706
conserved hypothetical protein XF1231 [imported] - *Xylella fastidiosa* (strain 9a5c)
C:Species: *Xylella fastidiosa*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: G82706
R:Anonymous. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis of Plant Pathogens
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below

A: Molecule type: DNA
A: Residues: 1-617 <SIM>
A: Cross-references: GB:AE003957; GB:AE003849; NID:9106207; PIDN:MAP8041.1; GSPDB:GNP8041.1
A: Experimental source: strain 9a5c
R: Simpson, A.J.G.; Reimisch, F.C.; Artuda, P.; Abreu, F.A.; Acenion, M.; Alvarenga, R.; Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreras-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A: Authors: Ferreira, V.C.A.; Perro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr J.D.; Jungueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins


```
OY 464 IETVELTDCILMDISPIEFSASNLIOQKLNLYAAKARHLIDMPDDRVLAINHDDGV----- 519
| : : : |
Db 284 --TIEVSEG-----KHRYFGVGAQY-----STTEGIGLOG 311
OY 520 ---NRSILG-----RISDAVASARAILPDESENEVIDLPE-----RTA 555
| : : : |
Db 312 YMGHRLNFGQAESLRIEGSVSRIAEA-----SSVEGMDYSAGITFTKPGMPNRTTFKTS 366
OY 556 LANRKTADVYQSKVPLVYVVASDKPRDQIGLGWSDGTRLVTKFEHNLINRDGYOA 615
| : : : |
Db 367 LIATKENDPTVRAKTL-----TGT---AGFAVELNDTDTAA 400
OY 616 GAELRL--SEDKKGVKLYATKPLSHPL-----NDQLRATIGYQOEFGHSTNGFDLST 666
| : : : |
Db 401 GLEVOAMDTEDAFGKNEYLT--TSIPLEFVRDTRDKLNPTGEGFRAS----- 445
OY 667 RTLEHEISRSIIIONGWNRTYSLARYLRLDKLKTQAPPETWQDLPVDFVNGKPSQEAALLAGV 726
| : : : |
Db 446 -----LAKPSYEAL----- 455
OY 727 AVHKTVADNLVNPARGYRQRYTSLEVSSGLVSDANMAIARAGISGVYSGDNAYGSNRH 786
| : : : |
Db 456 --NGTFFSSFECSITGYK-----GLGAEDRLIM-----AG 483
OY 787 QMTGIGAGYIWSDNFNHVPYRLRFPAAGDOSIRGYAHDLSPISDKGYLTGGQVLA VGT 846
| : : : |
Db 484 KLSGGVLVG--GSDLDIPTTRRFFAGGGSVRGYSVOEISPYNAAGADATGGRSVYVGS 540
OY 847 AEYNEFMKDLRLAVFGDIGNAYDKGFT---NDTKIGAGVGRMASPVGOVRVAVATGVK 903
| : : : |
Db 541 VEARIKVTDTIGLVFPFDAGVSD-GVTPPDSIDRAGAGIGLRATPGRPLRDVAMPLE 599
OY 904 --EEGNPKIKLHFITGTPF 919
| : : : |
Db 600 KYDGGNFGIYAGIGQSF 617

RESULT 13
H97672
hypothetical protein AGR_C_4742 [Imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: H97672
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: H97672
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-641 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK8337.1; PID:g15157817; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_4742
A:Map position: circular chromosome

Query Match 5.0%; Score 236.5; DB 2: Length 641;
Best Local Similarity 20.3%; Pred. No. 9.3e-06;
Matches 162; Conservative 107; Mismatches 246; Indels 283; Gaps 31;

OY 200 GSHQKTEPRANKALEDTTQESAMLNST-PRLRQTLAVARAVGYDIDLSIRNS 258
| : : : |
Db 49 GSEPEVEVINPKYA--VTLDA--DADKSLKSSLENSILLADKDKPASGDLGLIKA 104
OY 259 ICEVDYIIDLGEPRV-YIDYRAVEVRGEGADKFAFTVADE--VPLI---GDVFNHG 310
| : : : |
Db 105 RODRDLIALIYENARKGGIVNVTAGKANVDDLPRNPVDFHSTFVFWITVTTPGPKFTLG 164
OY 311 KYE---TKKNLIENASAEHGYFDGKRWMLDRSYDVILPD----- 344
| : : : |
Db 165 NVRLGSDVTGRLNDEYGLIAGDGAGSLAIRACGNKLLIDLAKAGRPPLAKLTKREAVANHA 224
```

```
OY 345 --NTADVSLITDGTQYRFDEVEVFTIDPKTNLTDPDKLPYKRELEOLLTVNMGAVN 403
| : : : |
Db 225 TMTVTITMAEGGPAPPLGAVTV-----TGETYDGD-----FIRYSRLNGGEPYS 271
OY 404 LOAVBALNDLIATRYFNWVNTVEIFPEREQIONQVSEFOSSSSRTPEAQVDESTLEPV 463
| : : : |
Db 272 PEKLRKKAADRLRLQGVF-----SLLTKEAGTLAROGTIFL 307
OY 464 IETVELTDCILMDISPIEFSASNLIOQKLNLYAAKARHLIDMPDDRVLAINHDDGV----- 519
| : : : |
Db 308 --TIEVSEG-----KHRYFGVGAQY-----STTEGIGLOG 335
OY 520 ---NRSILG-----RISDAVASARAILPDESENEVIDLPE-----RTA 555
| : : : |
Db 356 YMGHRLNFGQAESLRIEGSVSRIAEA-----SSVEGMDYSAGITFTKPGMPNRTTFKTS 390
OY 556 LANRKTADVYQSKVPLVYVVASDKPRDQIGLGWSDGTRLVTKFEHNLINRDGYOA 615
| : : : |
Db 391 LIATKENDPTVRAKTL-----TGT---AGFAVELNDTDTAA 424
OY 616 GAELRL--SEDKKGVKLYATKPLSHPL-----NDQLRATIGYQOEFGHSTNGFDLST 666
| : : : |
Db 425 GLEVOAMDTEDAFGKNEYLT--TSIPLEFVRDTRDKLNPTGEGFRAS----- 469
OY 667 RTLEHEISRSIIIONGWNRTYSLARYLRLDKLKTQAPPETWQDLPVDFVNGKPSQEAALLAGV 726
| : : : |
Db 470 -----LAKPSYEAL----- 479
OY 727 AVHKTVADNLVNPARGYRQRYTSLEVSSGLVSDANMAIARAGISGVYSGDNAYGSNRH 786
| : : : |
Db 480 --NGTFFSSFECSITGYK-----GLGAEDRLIM-----AG 507
OY 787 QMTGIGAGYIWSDNFNHVPYRLRFPAAGDOSIRGYAHDLSPISDKGYLTGGQVLA VGT 846
| : : : |
Db 508 KLSGGVLVG--GSDLDIPTTRRFFAGGGSVRGYSVOEISPYNAAGADATGGRSVYVGS 564
OY 847 AEYNEFMKDLRLAVFGDIGNAYDKGFT---NDTKIGAGVGRMASPVGOVRVAVATGVK 903
| : : : |
Db 565 VEARIKVTDTIGLVFPFDAGVSD-GVTPPDSIDRAGAGIGLRATPGRPLRDVAMPLE 623
OY 904 --EEGNPKIKLHFITGTPF 919
| : : : |
Db 624 KYDGGNFGIYAGIGQSF 641

RESULT 14
D97527
omp1 protein precursor (U51683) [Imported] - Agrobacterium tumefaciens (strain C58, C
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: D97527
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: D97527
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87173.1; PID:g15156447; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2554
A:Map position: circular chromosome

Query Match 4.9%; Score 232.5; DB 2: Length 774;
Best Local Similarity 21.0%; Pred. No. 2.2e-05;
Matches 178; Conservative 109; Mismatches 310; Indels 251; Gaps 39;

OY 165 IKRLYARLF-----NDGVNRY---PLKAKFYQSSQSGETSAIGSS 202
| : : : |
Db 77 VKRLIATGIFSNVSMRYSGLTLYVTVAENQLVNOYVFNGNRKIKDDKLGIATVOTQPMKPF 136
```


QY 203 HQK--TEPYAIKAKALEDTQESAMDNGSIPLRLOTALVAARAGVYIDLSII----- 255
 Db 137 NQAIYVADIAIRIKKAYSAGR-----SDVEITQTQV-----SVGGGRVNIAFVINEGE 184
 QY 256 RNSIGEVNIIHHDGEPYIDR-AVEYRGEGADKAFITVADEVPLIJDYFHHGKET 314
 Db 185 RTKIGRIDFI-----GNNSTSDGRLAIVNTKKSNMISFLTRK-----DVTYNEKDLRA 232
 QY 315 KKNLIENASAEHGYPDGRMLDRSVNIIIPD--NTADVSLIYDTQYRFDEVEVFTIDPK 372
 Db 233 DEEALRQFYRNGYADFRVY--SSDAVIDESKNEVYIISTVDGSKRTYFQGNVAVESTVP- 289
 QY 373 TNOITTPDKLPVRELLEQLLITVMGEAYNLDQAVRALSNDLIATRYFNKNVTEIYFPER 432
 Db 290 -----GVGSESLQGLVETRGQASVSAKEVQO-SMEALSKR----- 323
 QY 433 EQIONDVSEFOSSSSPREPAQVDESTLEPIETEVELTIDGILMDISPIERSANLIODKL 492
 Db 324 -----VAGEGYFPARTPPGRDGRMSGNTIGVYIYIDGGERAYVRITETRGTRFTDYV 376
 QY 493 NLVAAKARHLVDMDDRVLAINHDDGVNRSI-----LGRISDAVASAVARAILPDE 542
 Db 377 -----IRREED-----ISEGDAPNQITIAAKRRLREALGYFSKVNISTAGGSAPD- 421
 QY 543 SENEVIDLPERTALANKRTPADVYOSKVPPLYVPAASDKPRQOIGLGWGSDTGRTLVTK 602
 Db 422 -----RVYIVVDVEDOSTGSFGIGAGYSQNDGVILLEAS 454
 QY 603 FEH-NLINRDGY--QAGAEILRLSEDKGVKLYATKPLSHPLNDQIRATLGYO----- 651
 Db 455 VEERKNFLGRQOYINVAAG--EDDARTYISLSTFEPY-----FLGRLAAGFDL 501
 QY 652 -----QEVFGHSTNGFEDLSTRTLEHETSRSIIQN-----GGMNRTY 687
 Db 502 FKNQSKSEDDYNNYDEGFAL-----RYTAPITENLSPTTFKYTKQIYVEGSGDMQNNNA 554
 QY 688 SLTRYRDLKLTQAPPEITWQDLPVDEVNCKPQSOEALLAGYAAVHYKYADNLVNPGRGYRQRY 747
 Db 555 NL-----AEPQAL-----IRGEDWTQISLISNTLNNTLIDDRNM-PREGQOAL 597
 QY 748 SLEVGSSGLVSDANMALARAGISGVYSFGD--NAVGSNRAHQMTGIGIAGYISDQPNHY 805
 Db 598 TNEF--AGLGSDSYXXIYIAARYYTILSDEYDVG-----LTG--QAGHWPPTGDNL 648
 QY 806 PYRLREFPAGGDSIRGYAHDLSLPI--SDKGYLTGGOVYLAVGTAEXN-----YEFMKD 856
 Db 649 VFD-QEKFQGRQ--VRCEKNDIGIPRIGSDS--IGTTYFAFASAEVTAPMPGPEPF--G 701
 QY 857 LRLAVFGDIGNAVDKGT-----NDTKIGAGYGVNRASVGVQRYVDVAVGKKEGPN 908
 Db 702 LRLAVFGDAGTMTGNKYSTQTVKRDNSIRASAGIGVMASPFQPIRVYAIPIAKEDYD 761
 QY 909 IKLHFFIG 916
 Db 762 EEQRRFRFG 769
 RESULT 15
 AE2746
 group 1 outer membrane protein precursor omp1 [imported] - Agrobacterium tumefaciens (str. C58)
 C.Species: Agrobacterium tumefaciens
 C.Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 C.Accession: AE2746
 R.Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H.;
 erage, G.; Gillet, W.; Grant, C.; Guentherer, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan,
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 ;Reference number: AB2577; PMID:11743193
 ;Accession: AE2746

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <KUR>
A:Cross-references: GB:AE0086a8; PIDN:AA142387.1; PID:g17739796; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: omp1
A:Map position: circular chromosome

Query Match	4.9%	Score 232.5	DB 2	Length 774
Best Local Similarity	21.0%	Pred. No. 2.2e-05		
Matches 1/8: Conservative	109	Mismatches 310	Indels 251	Gaps 39

QY 165 IKRLYARLE-----NDGVNKV---PRLKAFYQSSQSGETSAIGSS 202

Db 77 VKRLYATGYFSNVSMRVSGSTLVVTNENQLVNVFNGNRRIKDKDKLAGIVQTQPMGPF 136

QY 203 HOK--TERYANIKALEDITQESAMDINGSIPLRQOTALVARAVGYDIDLSII----- 255

Db 137 NQAI VTAD IARI KEAY SAIGR-----SDVE ITTQT V-----SVGQGRVNIAFVINEGE 184

QY 256 RNSIGEVDI IHD LGEPVYIDYR-AVEVRGEGADDKAFTTVADEVPLIGDVFHGKYET 314

```

Db      185  RTKIGRIDFI---GNNSSDGRLAAVINTKKSNNLSFLTRK-----DVYNEDKLRA 232

```

QY 315 KKNLIENASAEHGYFDGRWLD RSV D V I L P D - - N T A D V S L I Y D T G T Q Y R F E D E V V F F T I D P K 372

Db 233 DEEALRQFYNRGYADFRV--SSDAVLDESKNEYTISITVDEGKKYDEGNVAVESTVP- 289

QY 373 TNQLTDPDKLPVKRELLQLLTVNMGAYNLQAVRALSNDLIATRYFMVNTEIVFER 432

```

Db 290 -----GVDSELGVLVETRQGSYSAKEVQQ-SMEAISKR----- 323

```

QY 433 EQIQNDQVSFEQSSSRTEPAQVDESTLEPVIEVVELTDGILMDISPIEFSASNLIQDKL 492

Db 324 -----VAGEGYPFARVTPRGDRMSGNTIGVTYIVDQGERAYVERIEIRGNTRTDYV 376

QY 493 NLVAAKARHLYDMPDDRVLAINHDDGVNRSI-----LGRISDAVSAVARAIIIPDE 542

```

Db 377 -----IRREFD-----ISEGDAFNQTIITAKRRLALGYFSKVNISTAGGAPD- 421

```

QY 543 SENEVIDLPERTALANRKT PADVYQSKVPLYVFVASDKPRDGQIGLGWGS DTGTRLVTK 602

Db 422 -----RVIVVDVEDQSTGSGFGIGAGYSQNDGVLLAS 454

603 FEH-NLRDGY---QAGAE LR LSEDKKGVKLYATKPLSHPLNDQLRATLGQ----- 651

Db 455 VEEKNFGRGQYIRVAAGAG---EDDARTYSLSTPEY-----FLGRIAGFDL 501

QY 652 -----QEVFGHSTNGFDLSTRLEHEISRSION-----GGMNRTY 687

Db 502 FKNQSKSE^{..}EDY^{..}NY^{..}DEQ^{..}GFAL^{..}-----RT^{..}APITEN^{..}LS^{..}TFK^{..}Y^{..}TY^{..}KQ^{..}IN^{..}YEG^{..}KG^{..}DW^{..}Q^{..}NNA 554

QY 688 SLRYYRLDKLKTQAPPETWQDLPVDFVNGKPSQEQALLAGVAVHKTVADNLVNPMPRGYRQRY 747

```

Db      555  NL-----AEPYQAL-----IRGEDWTQSI LSNTLNNTLDDRMN-PREGQAL 597

```

QY 748 SLEVGSSGLVDANMAIARAGISGVYSFGD--NAYGSNRAHQMTGGIQAGYIWSDNFNHV 805

Db 598 TNEF--AGLGDSYYKYAKARYYTLSD EYDVG-----LTG--QAGHVMPTGDNLL 648

QY 806 PYRLRFAGGDQIRGYAHDLSPI--SDKGYLTGGQVLAVGTAEYN-----YEFMKD 856

Db 649 VFD-QEKEGGRQ-VRGFKNDGIGPRIGSDS---IGGTTYFAASAEVTA¹MPGPEDF--G 701

QY 857 LRLAVFGDIGNAYDKGFT-----NDTKIGAGVGVRWASPVGQVRVDVATGKKEGPN 908

Db 702 LRLAGEVDAGTMYGNKVSTSTQTKDDNSIRASAGIGVMWASPEGPIRVDYAIPIAKEDYD 761

QY 909 IKLHFFIG 916

Db 762 EEQRFREG 769


```
Db 430 VVEKST-----GEFSGIGGYTGGESPGAQVEAAT 460
Oy 602 KEFHNIINRDGY---OAGAEILSEDEKKGVKLYATKPL-----SHPLND- 642
Db 461 --ERNLGNQYIRISAGAC---ODMKRNYGLFTEPEYFGLYSAGFDVFRSRVYND 515
Oy 643 ----QLRATLGYOOEVEFGHSTNG-----FDLSTRLEHEISRSII---ONGGWN- 685
Db 516 YVEEQTGCTIRFGLPITDNFSAGIAYSLVQEKYDLFRGDAENYAPALLAENSPLRS 575
Oy 686 --TYSLR-RDLKLTQAPETWODLPVPFNKPSOE-ALLAGVA--VHKYVADNLVNP 739
Db 576 SVSYSLYSIDIKN-----PHDGLYKGFIOEFAGLGDGDKKYVTKTEFGN---- 621
Oy 740 MNGYRORYLEVGSSGLVSDANNAIARAGISGYSFSGDNAYGSNRHQMGTGQAGTWS 799
Db 622 ----YYOTLQOADIYCLLG-----VGAGYIHERGDD-----GVRI----- 653
Oy 800 DNEHNYPLRFFAGGDQIRGYAHDSLSPISD---KGYLGGQVLAAGTAENYEF-- 853
Db 654 -----FDLFKNSSDIIRGFKNFGIGPYODAKNGKRYMGTTFYFSGTAEOFPMPV 704
Oy 854 ----MKDLRLAVGEGDIGNATDKGFTNDT-----KIGAGVGV--RWASFVQGVRVAVAT 900
Db 705 LPESLGVKRAFPADAATLYG---NDTPDISGDDKRLRASVGSLMWASPFGLRFDYAF 760
Oy 901 GV-KEEGNPIK-LHFFIGTFP 919
Db 761 PVAKADTKVQVNFNGVSTKF 781
```

```
RESULT 18
A82025
Probable outer membrane protein NMA0296 [imported] - Neisseria meningitidis (strain 2249
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: A82025
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: A82025
A:Molecule type: DNA
A:Status: Preliminary
A:Residues: 1-615 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:q7378778; PIDN:CAB83602.1; PID:q737905
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: NMA0296
```

```
Query Match 4.7%; Score 220; DB 2; Length 615;
Best Local Similarity 22.3%; Pred. No. 7.6e-05;
Matches 131; Conservative 72; Mismatches 236; Indels 148; Gaps 19;
```

```
Oy 462 PVIEIVELTGLMDISPIEFASNLIDKLVAKARHLVDMDDRLAINHDDGVNR 521
Db 48 POTESVKLKPFPVRIODSEIKDMVEHPLITQOOEEVLDKQGTGFLAEADPNVKT 107
Oy 522 SI--LGRISDAVSA-----VARAILPESENEVIDPRTALAN 558
Db 108 MLRSKGYSSKYSLEKDGAYVHTTGPRTKRIANVAVAILGLISDGLAEYNNALEN 167
Oy 559 RKTPA-----DYQSKR-----VPLVYVASDKP 582
Db 168 WQDPVSGDSDODSMNSKTSYLGAVTRKAYPLAKGNTRAAVNPDTATADLNVVVDSGRP 227
Oy 583 RDQIGLQWGSDTGR-----LVTKFEHNLINRGYQAGEL- 619
Db 228 ----IAGDEFTITGTOKRPEQIVSGIARFQETPDDLLDLDQAL-EQNGHSGASVQ 282
Oy 620 ----RLSEDKKGVKLYATKPLSHPLNDQLRAT---LG-----YQOEVFEGHSTNG 661
```

```
Db 283 ADFRLOGDRVPYKVSUTEVKBHKLKLENGIRLDSSEYGLGKIAYDYNLKFNGYIGSVYMD 342
Oy 662 FDLSTRLEHEISRSIIQNGW---NRTYSLRRDLKLTQAPETWODLPVDFVNGKPS 718
Db 343 MDRYETTLAAGTISOPRNYRGNNTSVNSYSTQNLEKRAFSGGIYVDRAGIDARLG 402
Oy 719 QEALLAGVAV-----HKT-----VADNLVNMRCGRQYSLSEVGSSGLVSD 759
Db 403 AEFLAECRKLPISDIDLGNSHATMLTASMKRQLNLNVLENGHYLDGKTGLTGLAEFLS 462
Oy 760 ANN--AARAGISGYVSFGDNAYGSNRHQMGTGI---OAGY-IWSDNENHYPLRFF 812
Db 463 TALIRTSARAG-----YFFPENKKLTGFIIRGQAGTVADNMN-VPSGLMFR 510
Oy 813 AGEDQSRGYAHDSLSPISDKGYLTGGQVLAAGTAENYEFMDRLAVFGDIGMAYDKG 872
Db 511 SGGASSVRYGELDSIGLAGPNSVLPERRALLVGSLEYQLPFTRTLSCGAVFHDMDGAANF 570
Oy 873 FTNDTKIGAGVRYMASPVQGVAVDVAATGVKEGEGNPIKLFHFFIGTFP 919
Db 571 KRMKLHSGSLGVWFSPILAPFSFDIAYGHSDK-KIRWHISLGTFR 615
```

```
RESULT 19
A81430
Outer membrane protein Cj0129c [imported] - Campylobacter jejuni (strain NCCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: A81430
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chli
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; Vanlille, A.; Whitehead, S.; Bar
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: A81430
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-739 <PAR>
A:Cross-references: GB:AL139074; GB:AL111168; NID:q6967505; PIDN:CAB72613.1; PID:q696
A:Experimental source: serotype O2, strain NCCTC 11168
C:Genetics:
A:Gene: Cj0129c
```

```
Query Match 4.6%; Score 216.5; DB 2; Length 739;
Best Local Similarity 18.1%; Pred. No. 0.00016;
Matches 172; Conservative 142; Mismatches 317; Indels 317; Gaps 43;
```

```
Oy 68 ARLNMA-----GLNAKPGQALDVVNFDDQSPISRIGEGSPPLGLDMSVTEETPLIS 119
Db 13 AMANNAATTIDIKFIGLNLHLSNTSAINIAGL-----KIGEEINPAKIMTALIN---LY 61
Oy 120 LEELEFAOESTENGINPNDIYPEYGEOPN-SEYVVPPTLEPEKPGI-----TKRLARLE 173
Db 62 KKNYFENIAVE---NNNGLEIIVTEKPIYATYTTIGTASNDKQVESLGLTKR--GTL 116
Oy 174 NDGVNK--VPRKANFYOSS-----OSGETSAIGSSH--- 203
Db 117 DEGNIKAEIERIKAYEAKSYFDYVEYKKTLLENTDGLELEFYVNRGNIILIDNVHLSG 176
Oy 204 OKTEPYANIKALDEITQP-----SANDLNGSITRLKOTALTAAVARANGYDI 250
Db 177 AKKESYSIDIEPAVNVKKEFEFGMMGMRNDGKLKVELSLSDSSRIADVEW---KKGYL 232
Oy 251 DLS-----IIRNSIGEVDVITHDGEPEYIDYRAVEVREGADKAFPTVADEVPLIG 304
Db 223 QYSSPYLKTYYDTTYQANLTYFIKE-GKPKYI--KSISLENPLFDKONAKQYVKDLRSSAG 289
Oy 305 DVFHGKYETKKNLLENASAEHGYFDGRMLDRSDVILIPDNTADVSLYDTGTQYRFDEV 364
Db 290 KTIINIEDIRKDVKTETQESADIGY---AFVEYVPDIQKNDQGEATVYFKV---IPHKV 343
Oy 365 VFTTIDPRTNQLTTPDPLKPKRELLEQLLYVNGEAYINLQAVRLASNDLIATRTYFMNVN 424
```

```
Db 344 YNNIISGSRIVD---RVYRRELY-----ITEGNLYNRTDISESKNAKRTSYFDDVN 395
QY 425 TELVPERREQIQNDQVSEFGSSSRTEPPAQVDSSTLEPVEIETEL-----TD 471
Db 396 -----IKEEKVDOTHTDILYDVKEASGALISGGIGYSSD 430
QY 472 GILMDISPIEFASNLQDKLNLVAAKARHLVDMPPDRVLAINHDDGVNRSITGRISDAV 531
Db 431 GILLNAS---LSDPTNIFGSGI-----KSSVSVKSDSD-----TLSGRIS--- 466
QY 532 SAVARALPDESENEVYDLPERTALANKRTPADYQSKVPLVYFVASDKPR--DQIGL 589
Db 467 -----LVN-----PRVLDQSYSL 479
QY 590 GMSGDYTRLVTKFEHNLINRDGYQAGAEIRLSEDKKGVLYATKPLSHPLNDQLRATLG 649
Db 480 G-----GT-----LVS-----ND----- 487
QY 650 YQGEVGHSTNGFDLSTRLEHEISRTIIONGKNRTYSLRYRLDKLTKQAPPETWODLP 709
Db 488 YEMDNVSEKNYGFDI---TIGROFARY-----NVSLTYNLEQ----- 522
QY 710 VDPVNGCPSEALLAGVAVKTVADNLVNPMDRGYRQRYSLSEVGSSGLVSDANMAIARAGI 769
Db 523 SDYIHLSPT--LRTGYELKSKTKSS--ITPAITPNDTDYILPRSGIT--ASTSLEYAGL 577
QY 770 SGVYSEGDNAVGSNRHOMTGGIO--AGY-----IMSDNHNVPYRLRFEPAG 815
Db 578 GGDQEF-----ISSSKFNFYQLODYGIDYIRYKASFYKVMDEGY--LPINQRIYLG 631
QY 816 DQSTIRGAHNSLSPISKGLVLTGGQVLAAGTAENYEFMDLR--AVFGDIGANAYRKG 873
Db 632 TRSIRGESRTVSPKQWQGEIGETIAFANSVELSPLIDRIKIRGSGVPDYGMGRK 691
QY 874 TNDTKIGAGVGMASFPVGOVRVAVATGV--KEEGNPRIKIHFFIGTPE 919
Db 692 DEIKRMSTGICIEIMPITGIPLOLVFAKPLMDKKDDTNSEFNCTGRF 739
```

```
RESULT 20
B87448
conserved hypothetical protein CCI1603 [Imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: B87448
R:Letman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: B87448
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-628 <STO>
A:Cross-references: GB:AE005673; NID:g13423000; PIDN:AAK23582.1; GSPDB:IGN0148
C:Genetics:
A:Gene: CCI1603
```

```
Query Match 4.6%; Score 216; DB 2; Length 628;
Best Local Similarity 21.8%; Pred. No. 0.00013;
Matches 110; Conservative 87; Mismatches 236; Indels 72; Gaps 17;
QY 456 DESTLEPVETVELTGDILMDISPIEFASNLQDKLNL---VAAKARHL-----D 504
Db 155 DEGVROAAVAMRLTEGEPSADYVGAEGRIYAQVAKLGVADVAAPREVYVDHADRTV 214
QY 505 MPDORVLA-----INHDDGV-----NRSITGRISDAVSAVARALPDESENEVYDLPERT 554
Db 215 RPFPRIMAGLVRLNGVDVYTKGRTNPEWVGRLAPWYA-----GDVYDPEVYA 262
QY 555 ALANRKTADVYOSKVV-----PLYVVASDKPRDQIGLQSGSDYTRLV 600
```

```
Db 263 ELERRLDATAVYDYSISLACTDKASAEGRVAVVTLSDRRARITELGAGTISEGACVD 322
QY 601 TKFEHNLINRDGYQAGAEIRLSEDKKGVLYATKPLSHPLNDQLRATLGYYQGEVGHSTN 660
Db 323 ARWR--YNRCKRADITTYALRFAKLEORLGAELISLPHWRPQ--QTLKLNSSVFRNTD 378
QY 661 GFDISTRTLEHEISRTIIONGKNRTYSLRYRLDKLTKQAPPETWODLPVFNKPSQE 720
Db 379 AYNETGATVGVDLTR--RQTTAYRTGVSPDLSQTEQVVRN-----GLINGRKLN 429
QY 721 ALLAGVAVHK--TVADNLVNPMDRGYRQRYSLSEVGSSGLVSDANMAIAR--AGISGYSEGDN 778
Db 430 ATTLGLAAVAMDSDDIILDPKRGKLETRAE--PTYAGDTSVYRLKLAGGSAYL----- 483
QY 779 AYGSNRHOMTGGIOAGYIWSDNHNVPYRLRFEPAGDOSIRGYAHDLSPISEKGYLTG 838
Db 484 PFGKQDSTVLAARVKLAILGAGLLDVPASRRPFGSGGSVRYAYQAIGPRLSDNFPQG 543
QY 839 GOVLAAGTAENYEFMDRLAVFGDIG--NAYDKGFINDTKIGAGVGMASFPVGOVRV 896
Db 544 GISLVETSFVEYRQKITDRWSGVAFVDAGATGTHETPOREDFRAGAGLVRYDLGFGPIRA 603
QY 897 DVATGV--KEEGNP--IKIHFFIGTPE 919
Db 604 DIAPLGRKKDPRKFQYLSIGQSF 628
```

```
RESULT 21
G81003
conserved hypothetical protein NMB2134 [Imported] - Neisseria meningitidis (strain MC
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: G81003
R:etzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qiu, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Piazza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: G81003
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-635 <STO>
A:Cross-references: GB:AE002561; GB:AE002098; NID:g7227384; PIDN:AAK42442.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB2134
```

```
Query Match 4.5%; Score 215; DB 2; Length 635;
Best Local Similarity 22.1%; Pred. No. 0.00016;
Matches 132; Conservative 73; Mismatches 225; Indels 166; Gaps 20;
QY 462 PVIEVETVELTGDILMDISPIEFASNLQDKLNLVAAKARHLVDMPPDRVLAINHDDGVNR 521
Db 68 PDTESVKAKRPFVPLIDQDESEIKDWEHPLITQOGEVLDEKQGFLEAPADNVKT 127
QY 522 SI--LGRISDAVSA-----VARAILPDESENEVYDLPERTALAN 558
Db 128 MLRSKGYFSSKVSLEKDGAYVHTTGPRTKIANVGAVALIGDILSDNLAEYRNALN 187
QY 559 RKTPA-----DVYOSK-----VPLVYFVASDKP 582
Db 188 WQPFVSGDFDQDSWENKTSVLGAVTKAIFLAKLGNQAAVNPDTATADLNVYVDSGRP 247
QY 583 RDQIGLQSGSDYTR-----LVTKFEHNLINRDGYQAGAE 619
Db 248 ----IAGDFEITGTQRPQEQIVSGLARFGPMPYDLDLDFQAL--EQNGHYSGASVQ 302
QY 620 ----RLSDKKGVLYATKPLSHPLNDQLRAT--LG-----YQGEVGHSTNG 661
Db 303 ADFRLQGDVRAVAVSVYEVKRHRKLEGTGIRLDSEYGLGKIAVDYVNLFNKGYSVYVD 362
```

```

0Y 662 FDLSTLTHLHHSIIIONGWM---NRTYSLRKLDKLTQAPPEPWODLPVDFWNGKPS 718
      || || || || || || || || || || || || || || || || || || || || ||
Db 363 MDYETTTLAAGISOPRNYRCGYWMTSNVSYNRSTONLEKRAESGQW-----YVRDRAG 416
      || || || || || || || || || || || || || || || || || || || || ||
0Y 719 QEALLA-----GVAV-----HRT-----VAADLVNPMFGY-----RQRY 747
      || || || || || || || || || || || || || || || || || || || || ||
Db 417 IDARLGAELAEGRKPIGSAVDLGNSHATMLTASMKROLLNNVTLHENGHYLDCKICTTL 476
      || || || || || || || || || || || || || || || || || || || || ||
0Y 748 SLEVGSSGLVSDANMALARAGISGYVSFGDNAYGSNRAHQMTGT----OAGYIWSDNFN 803
      || || || || || || || || || || || || || || || || || || || || ||
Db 477 GFLLSTALRTS---ARAG-----YEFTPENKKLGTFIIRQAGYTVARDNA 521
      || || || || || || || || || || || || || || || || || || || || ||
0Y 804 HVPYRLREFAGDQSIKRGYADHSLSPISDKGYLTGGQVAVYAEYNEFMRKDLRLAVFG 863
      || || || || || || || || || || || || || || || || || || || || ||
Db 522 DVGSGIMFRSGGASVRYGELDLSIGLAGPNCSVLEPERALLVGSLEYQLPPTRLTSAGVFH 581
      || || || || || || || || || || || || || || || || || || || || ||
0Y 864 DIGNAVDKGTNTDKIAGVGYRNASPYGVQVRVDYATGVYKEGCPNPKLHFETGTPF 919
      || || || || || || || || || || || || || || || || || || || || ||
Db 582 DMGDAAANERKMKLKHSGSLGVRWFSPFLAPESFIDAYGSDK--KLRWHISLSTRF 635
      || || || || || || || || || || || || || || || || || || || || ||

```

RESULT 22
 E82731
 outer membrane antigen XF1046 [imported] - *Xylella fastidiosa* (strain 9a5c)
 C:Species: *Xylella fastidiosa*
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: E82731
 R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequences
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A55328 below
 A:Accession: E82731
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-784 C:STIM>
 A:Cross-references: GB:AE003941; GB:AE003849; NID:g9105978; PIDN:AAF83856.1; GSPDB:GN001
 R:Stimpson, A.J.G.; Relnach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carter, H
 as-Neto, E.; Docena, C.; El-Dorri, H.; Fachinani, A.P.; Ferreira, A.J.S.
 Submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitaajima, J.P.; Klieger, J.E.; Kuramae, E.E.; Laig
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
 A:Authors: Martins, E.M.F.; Matsushima, A.Y.; Menck, C.F.M.; Miranda, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.;
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A55328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF1046
 A:Superfamily: protective surface antigen D-15

	Query Match	4.58;	Score 214;	DB 2:	Length 784;	
	Best Local Similarity	20.58;	Pred. No.	0.00025;		
	Matches 177;	Conservative 120;	Mismatches 300;	Indels 266;	Gaps 45;	
QY	228	NGSIFRLPRTALVAARAVGYDDIDLSIIRNSIGSEVDVLIHDLGEPVIDYDRAVEVNGEGA	287			
		:	:	:	:	
Dd	17	NFSLVVLITDAA-----ESFEVANDI-RVDG-LQRIASGVFTFLPVN-RGDIV	60			
			:	:	:	
QY	288	DDKAETTVADVEPYLLIGDYVFHHG-----KYETR	315			
			:	:	:	
Dd	61	DD--AKVADAIRALKRTGFENVRIDRGNIILVKKERRAINKLTITGNKDISEELL	117			
			:	:	:	
QY	316	KNLINASAEHEGYFDGRWMLDRSDVLIIPDNTADVSLIYDTGTQYRFEDEVVEFTIDPKTNQ	375			
			:	:	:	
Dd	118	KGLSIGLSEGGEFFDLSDRV-----TQELKROYNNGRKVNV-----O	156			
			:	:	:	

Oy	376	LTTPDDKLPVYRELLEJOLLTYNMGEAYNLQAVRALSNOLIAFRYNNMNTLVEYPERRO-	434
Db	137	MTT--TTTPBLOKRNVDVTAIKEGKAARIH----NLIGKEKRNKDVSAMSEKSHN	209
Oy	435	-----IOMDOVS-----FEOSSSRTEPAQVD-----ESTLEP-----YLETVELDGG	472
Db	210	WASWRRDDQYSKEKLSGDELEKLNWYLDRCGYVDNDISTOVSISEPKHNFITAGYIEG	269
Oy	473	ILMDISPIEFESANLI-ODKL-NLVAARAKRLYDMPDDRVLAINHDDGVNRSILGRISDA	530
Db	270	DQYKISSIKVGNVFLPQEKITEKLVIPKTDIF--SRVL-LEVSAAIINTLSNIGYA	324
Oy	531	VSAAARALLPDESENEVIDLEPRTALANRKPIPADYQSKKPYL--YVVASDCKPRDGOIG	588
Db	335	FSKVN----PIPTANRA----DRYAVN----LHYIPGRYVRQILFKGMTRTSD----	368
Oy	589	LGWGSDDTGRVLTKEEHN-----LINRGYQAGAEHLR-----SEDKKGVLY	631
Db	369	-----EVLRRREKROEKNWSYSOAALDRSKIRLQRLGYEAVDVESTPYGSDNDQYDI--VY	422
Oy	632	ATKPLSHPLNDOLRATILSYQOEVFEGHST-----NGFPLSTRLEHEISRTLIQ-----	679
Db	423	TYKETT---SGSFQVGLGY-SKTYGVTTISVOLSQNNFLGSGNRVSVDSRSRYDRIYSFS	478
Oy	680	-----NG---GNRRTY-----SLRYRLDKLKTQ---APPETMODLPPVDFVNGK	716
Db	479	YTNPFETDNGVSLGYNLAYOKUDYSDFNAAQYNSKRMSGOTIFGIPITEND--TVSMVIGA	537
Oy	717	PSQELLLAGVAVHKTVAD-----SLRYRLDKLKTQ---APPETMODLPPVDFVNGK	751
Db	538	DSNOJTTTPPGSTPKAIIYIDAVGQRTFRAMRTELGMARDRTRNDYFPMNLGMYOIGAEV	597
Oy	752	---GSS-----GLVSDAMMAIARAGISGYVSGDMAYSNRAHQWGTG	790
Db	598	TLPGSTIYYKINQIOISKYMPILPALVLTNRLY-----GYGDD--YKSSHTRIILPD	647
Oy	791	GIQAGYIWSDNFNHVPYRLREFPAGGDOISIRGYAHDSLSPISD-----KGYLTGQVAV	844
Db	648	GTVA-----TASGLPFEENFYAGCTNSVRRFRONTLGLPBEVATLYNQGGPPLGSPFTV	701
Oy	845	GTAETNYEFMKD---LRLAVFGDIGNAVD--KGF--TNDFKIGAGYGVWMASPVGOVRVY	898
Db	702	GSTEMYEFKLPDSSARISAFALDFCGWENGVNNEKANELRASSGVALLMRAPIGIPISY	761
Oy	899	ATGVKEEBCNP--IKLHPFICITPF	919
Db	762	APPIKKNNDIEIRLOFTFGQOF	784

RESULT 23
AC0129
Probable surface antigen YP01052 [imported] - *Yersinia pestis* (strain CO92)
C:Species: *Yersinia pestis*
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AC0129
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titchell, R.W.; Holden, M.T.G.; Prentice, M.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barral
Nature 413, 523-527, 2001
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0129
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-795 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89894.1; PID:g15979119; GSPDB:GN00175
C:Genetics:
A:Gene: YP01052
C:Superfamily: protective surface antigen D-15

Query Match 4.58; Score 211.5; DB 2; Length 795;
Best Local Similarity 19.18; Pred. NO. 0.00036;
Matches 177; Conservative 139; Mismatches 306; Indels 283; Gaps 42;


```

Oy 542 ESENEVYIDL-----ERTALANKRTPADVYQSKKVL-----YVFASDKPRCOI-- 567
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 365 VSENGVTLQVAGGVENISVFRFRNKEGQDVNEOQPIRGRTQDIITREVELKPCOVFN 424
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 588 -----GLGMSDGTGR 598
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 425 RNTVQKDLQVRFTGGLFEDYAVSLDPGDPTRKVVNVVNVVENSQSIAGAGISSSGLE 484
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 599 LVTKFEHNLINRDGYQAGAEIRLSEDKKVKRLYATKP-----LSHPLNDQLRATL-- 648
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 485 GWVSYQQOONLGNRQKLGAEVQLGERELFLDRFDPMIGGDPRTSYTANIFRRRSISL 544
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 649 -----GYOOEV-----FCHSTNGFDLS--TR-----TLEHISRSLIIONGMWRTYLLRR 692
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 545 ITDGKDEDITFPDPGNPDNTNGODPRVRYRLGGVTFTRPPLSANPERAEWIASAGLOYO 604
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 693 LDKLKTQAPP-----ETWQDLVPDFVFNKPSQEALAGVAVKTVADNLV 737
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 605 --RVSTRDADGNGLRKQDAVFDGDDNGNRRTSSEIPLSF--SGTGEDDLLVQLGADRDLANNPL 661
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 738 NPMKGYRQRYSLSEYSGGLYSDANMAIARAGISGVY-----FGDNAGVSNRAHOM 788
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 662 QPFGSGFLRRG--VDOSVPYSGNIFLTR--FRGYSOYLPPYKFTGF-----SKGPETI 711
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 789 TGGIAGYIMSDNHNHYRFLRFAPGAGDOSIRGVASHDSLSPISDGYLGTGGOVLAVGTAE 848
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 712 ARIHQGYIVGD-----LPPIEAFTLGGSNSVRYG-----EGCALSSGHSFVQASVE 758
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 849 YNVEFMKDLRLAVFGDIGMAYDKGFTNDTK-----IGAGVGVRRMASPYGVQVREDA 899
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 759 YRFPPVSYSVGALEFFDVG--DLGTSRTALVLNKSGSGYGLGVRVQSPIGPINDY- 815
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 900 TGVEKEGNPIKLHFFICTPF 919
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 816 -GINDGDS-RINFGIGERF 833
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 29
F71910
  Probable outer membrane protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variatey: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: F71910
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
  Ives, C.; Gibson, R.; Metberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
  Nature 357, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
  A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: F71910
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-906 <ARN>
A:Cross-References: GB:AE001492; GB:AE001439; NID:g4155147; PIDN:AAD06181.1; PID:g415515
A:Experimental source: strain J99
A:Genetics:
A:Gene: jhp0600

```

	Query Match	4.18;	Score 194.5;	DB 2;	Length 906;	
	Best Local Similarity	19.9%;	Pred. No. 0.0041;			
	Matches 192;	Conservative 131;	Mismatches 311;	Indels 333;	Gaps 49;	
OY	116	TPPLSTLEELFAOEST-----EMGINPNDIYP-EYQ-----GEOPNSEVVV	153			
		:: : :	:: : :			
Db	37	TREKAKNEAQNETSOSNQPKKKVKISIVGLSYMSDMLANETAKIRVGMDVDSEKKID	96			
		:: : :	:: : :			
OY	154	PTLEPEKPGILIKRLYLARLFNDGVNKNVPRLAKAFYOSSQS-ETSAIGSSHQKTEPRYN	211			
		:: : :	:: : :			
Db	97	FVLALFNCGYFADYYA-TFENGJ-----LEFHDEKARIAGVEIUKGYTKEKEDOLSKQ	150			
		:: : :	:: : :			
OY	212	IKAALEDIQESAMDLDNGSLPRLKOTAL-VARAVGYVIDLSTRNSIGEVDVIITHDLS	270			
		:: : :	:: : :			
Db	151	MIKKKDGTDEQKLE-----NAKTALKTALEGOGY-----GSV-----	184			
		:: : :	:: : :			

QY	271	EEVYIDYRAVEVRGSGADDKATTYAAD---	EEPLLIGDVFNHGKQETKKNLIENASAH-	326
Db	185	-----VEETKEKVSSEGALLIVFDVNRGDSIYIKOSIT	EGSDKLRKRYTESLAKQ	235
QY	327	-----GYFGR-----	WLRSVD--VILPD-NTADVSLIYD	354
Db	236	RDFKMGMMGLNDGKRLRLDQLEYSRLIDQVYRRORYLDAH	ISSPLFKTDFTSHDAKLNHYK	295
QY	355	T--GTQYRFEDEVFEFTDPTKTNQLTTPDPKLPVKRELL	EOLLTVNMGEAYNLOAVALS	412
Db	296	VEKGIOYRISDIL-----	IEIDNPVPELK--TEKAKVKRKQVFNIEHLRAAQ	343
QY	413	DL-----IATFYEMVMTVEIYFPEREOIQNDQVSEB	SSSRTEPPAQVEST--LEPIETVE	468
Db	344	IKTELEADGVAFA--VYKPDLDK-----	DEKNGLVKTYIE	379
QY	469	LTDGLIMDSPLEFSASNLIDOKNLVAKARHLVDMPDR--	VLAINHDDGVRSILG	525
Db	380	VGD--MVHINDVIISGNQRTSDRI-----	IRRELLGPKDKKNLKLKLNSESLAR--LG	430
QY	526	RISDAVSAAVARAILPDESENEVYIDLPERTALANRKT	PPADYQSKVPLYVYASDKPPDG	585
Db	431	FEFSKVIIEKRY--NSSLMDL-----	LVSVEGRGTG	459
QY	586	QI--GLMGESDPTGRFLYTKFEHNLINRDQYOGAELR	SEDD--KKGVKLYATKP-----	635
Db	460	QLOFGLGYSYGLML---NCSYSERNLEFGTGSMSLY	ANATGCGGSHYRPMCPAGARM	515
QY	636	-----LSHPDLQDRLATLGLYQOEVFCHSTNGFD	STRLL--EHEISRIIONCG-----	682
Db	516	PAGNLSTNP-----	RIF--DSWYSTINLYADRYHISYQIQGGGCGFVN	558
QY	683	-----WNRFY--SLGRRLDKLTKQAPPETWQDLP	VPFVN--GKPSOELLACGYAVHK-----	730
Db	559	VGRMLGNRTHVSLGYNLWTKLGLFSSPLYNKYSSVNE	VASPRCSPASVITIRLSGG	618
QY	731	---TYADNLVNP-----	MRYROR-----YSLEVG-----SSGLVSDA	760
Db	619	KRPLPVESSSSSGAITTSPEIKGIMDRDXYHPT	ITSSEFLDYSYDNTDOYFPRNRVITS-	677
QY	761	NNAIIAAGISGYVSFGD---NAYGSNRAHOMTG-	-----IOAG	795
Db	678	-----SVATMSGSPSSGGLTNSMNGLGANVRNTKY	GKFAAYHHLQKYLIDLIARPKTGG	733
QY	796	YIMSDNF--NHVYRKLRFPAQSDSIRGAHDSLS	ISDKGYLTGGOVLAVGAELNYEFM	854
Db	734	YIFRNTDYLILNSTFYMGVTVYGRNGSITPKDEG	LMLGDDGJFTASTELSYVL	793
QY	855	K--DLRLAVFGDIG-----	---NAYDKGFINDT--KIDAGGV	885
Db	794	KAAKKRIAMFFPFGELTKPTKRGSPFYNAPTTAN	KPDYGVYAGGERATIRASTGLQI	853
QY	886	RMASPVG	892	
Db	854	EWISPMG	860	

RESULT 30
AE2417
hypothetical protein alr4893 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AE2417
R:Kaneoko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriugu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2417
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-676 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW6592.1; PID:q17134030; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr4893

Query Match 4.18; Score 193.5; DB 2; Length 676;
Best Local Similarity 20.18; Pred. No. 0.0029;
Matches 191; Conservative 119; Mismatches 323; Indels 317; Gaps 47;

```
OY 13 MPVALAAYLPMTSQAALAOONNPAIINHVPAHDATINAAKAGMPVLLPEQOARLNA 72
D 1 MRISASALLLTLLVA-----SNATQQAIAA-PHRASTPSK----- 35
OY 73 AGLNAKPOALDVNVDQSPISRIQSPPLGIDMSVIEETPLSLLELFAQESTEMG 132
D 36 -----QNEHL-VIIPVEETPV-QLCNVTPP-----ET-----EVTPOFSTSR 71
OY 133 INRDYIPERVOGEQPNSEVYVPPLEPEKGLRLKRLARLND-----GVNKKVRL 183
D 72 VIAQ-----NSPVVLPSPRPVPPAPKP--ATTVNELVATDVQIVGATPEL 118
OY 184 K---AKFYOSSQGETSAIGSHOKTEPYANIKALEDITQESAMDINGSIPLRQATALV 240
D 119 QELIRQVIRKOTGSDT-----QTQLQRVVAILE-----TGLF 152
OY 241 AARAGVYDIDLSIIRNSIGEVDYIHDLEPYYIDYRAVEVRGEADDKAFT-TVADDEV 299
D 153 ASANVN-----SRTTPSGILNVYO--VQPAIV--RSIQLTGA----KALTYSVAQ-- 194
OY 300 PLLIGDVFHHGKYETKKNL-----IENASAEHYFPGGRWL--DRSDVILLPMTADVS 351
D 195 PRQSOI---GKPISEGLQAOAAQVQWYADNGYNLARVLSTIEPNQGIININVAE--GL 250
OY 352 IYDTGQYRDEVEVFTIDRKTNQL--TTDPDKLPVKRELLEQLLTVMGEAVNLQAVRA 409
D 251 VSD--IKFRFVNDQKTIIDNGNPGVGRTPD-----FLRQQLKIQPGVQENITVKQ 301
OY 410 LSLDLITRFVNMVNTIETVPEREQIONDOVSFEQSSSRTEPAQVDESTLEPIETVEL 469
D 302 DYQQLRTGTFQSVN----- 316
OY 470 TDGILMDISPIEFSASNLIDKLLNLVAAKARHLXDMDDRVLAINHDDGVNRSI--LGRI 527
D 317 -----VAIAGD-----ATKIDMIELEKENGARAINLCGSTNGVGLMGTLL 356
OY 528 SDAVSAVARAILPDESENEVIDLPERTALANRKTTPADVQSKVPLVYFAASDKPRDGOI 587
D 357 N-----YODNIGGKNDTLLN-----V 374
OY 588 GLGMSDPTGRLVTKF--EHNILNRD--GYQAGAEHLRSEDKKGVKLYATKPLSLPLNDQ 643
D 375 GL---STTDQTFDTKTSPIRDTNSDRLYTVNA-----FRREISIEETDE 418
OY 644 LRATLIGQOEVEFHSTNGFPLSTRTLEHSRIISIIONGGNNRTYSLRYRLDKLTKTO--A 700
D 419 IRLANDQVRE--GKVGIGISLQ--RPID-----GNNASIGIFMYSKRTSTDRGNI 465
OY 701 PRPTMODLPVDVFNKPSQEQALLAGVAVHKTVDNLVNPGRYRORYSLEVSSGLVSDA 760
D 466 TPTDAGNPLS--VSGTGVDLTTVSFSATKQDRDNPIPTQGSVVAVSTE--GSVPICG 522
OY 761 NMAIARAGISGVSPGDMYAGSNRAHQMTGIGAGYIWSNPNVPRRLRLEFFAGDGSIR 820
D 523 NISMRLKADYSQYVYVNIIFNSQTPQVFAINVOAGTVLG--MLPYE--TFNIGGSNSVR 578
OY 821 GVAHDSLSPISDGYLVGGVLAAGTAAYNYEEMKDLRLAVFGDIGNAYDKGFT----- 874
D 579 GY-----DAGNVGSGRSYVLASAERYRPLVPIVGLFADPFAADLSGDTVLGNPA 629
OY 875 -----NDTKIAGVGYRMAKSPVQGVYDVATGVKKEGKNPLKLFHFTSTPP 919
D 630 GVNRKPGSGCYGAGIRVDSPLGLIRADY--GINDGES--RVHLIGIQORF 676
```

RESULT 31

outer membrane protein CC1915 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence.revision 20-Apr-2001 #text.change 20-Apr-2001

C:Accession: F87486

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87449; MUID:21173698; PMID:11259647

A:Accession: F87486

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-769 <STO>

A:Cross-references: GB:AE005673; NID:q13423368; PIDN:AAK23890.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC1915

Query Match 4.08; Score 190.5; DB 2; Length 769;
Best Local Similarity 20.08; Pred. No. 0.0053;
Matches 151; Conservative 94; Mismatches 250; Indels 259; Gaps 34;

```
OY 250 IDLSIIRNSIGEVDYI--IHDLEPYYI--DYRAVEVRGEADDKAFTTVADDEVPLLISDV 306
D 160 VDLVFEINEGAKSGVLGINFLGNMEYSDNDLRDVIIVTESRWYKILTS-----MDN 210
OY 307 FHHGKYETKKNLIEASAHEGYFPGGRWLIDRSYDYLPP--NADVSLIYDTGQYRFEDEV 365
D 211 YDPRIETEDRQLRKHYRNRYGFFERVIS--SVALPAPKNGPAVYTLTEEGPKYFGKI- 268
OY 366 FFTIDPKTNQTLTTPDKLPVKRELLEQLLTVMGEAVNLQAVRALNSDLIATRYENMYNT 425
D 269 --TYETELKTKLDGN-----LLAQILPRTGGLYE----- 295
OY 426 EIVPEREQIONDOVSFEQSSSS-----RTEPAQVDESTLEPIETVELTDLGILMDISPI 481
D 296 ---DERIEQATDALTFAAGAGAFVDVPRPYVPRNETKTVDVVFQYREG----- 342
OY 482 FSNASNLIDKLLNLVAAKARHLXDMDDRVLAINHDDGVNRSILGISDAVASAVARAILPD 541
D 343 ---PRVYDRIDIV--GNTRTL--DYVLRRELEVABGDAYNR-----VLVD 381
OY 542 ESEN-----EVID-----LPERTALANRKTTPADVQSKVPLVYFAASDKPRDGO 586
D 382 RSKNNMRLGFFKEVEIEDAPGAPDRSLR-----YKVEQPTGE 422
OY 587 IGLGMSDPTGRLVTKFEBHNLINRDGYQAGAEHLRSEDKKGVKLYATKPLSLPLNDQLRA 646
D 423 LSFSGYSSIDKLIV--LDVGITERNFRGQNLN-----ARA 457
OY 647 TLG--YQOEVFGHSTNGFPLSTRTLEHSRIISIIONGGNNRTYSLRYRLDKLTKQAPPET 704
D 458 SVGSLRQQLIDFGESEPRF-----LGRNLV--AGVN--LYTFRIYDLSEPAAYDTKSV 504
OY 705 WQDLPVDVFNKPSQEQALLAGVAVHKTVDNLVNPGRYRORYSLEVSSGLVSDA 760
D 505 GGDVRFEGFPLTNDSSMSLRVTRQDEVSVADSLCASSGSVQILCLRGAYITSLIGYGLR 564
OY 737 -----VNPGRYRORYSLEVSSGLVSDANMAIARAGISGVSPGDMYAGSNRA----- 785
D 565 IDKRNDPINPFRG-----FADLNQDLAGV-----GDVKKLKTQADAGW 604
OY 786 -----HOMTGIOAGYIWSNPNVPRRLRLEFFAGDGSIRGYAHDLSLP--ISDKG 834
D 605 YWGPTKLVYSATSSF--GYIEGNGGDNVRLNDRFYRGG--TSFGFELAGIPRPDISSSF 661
OY 835 YLVGGVLAAGTAAYNY--EPMKD--LRLAVFGDIGNAYDKGFTNDT----- 877
D 662 NSMGAKKLYAISTFELVPTFLEPQYGIKAALFSDVGTG--GLLDVDYDRGRSPGVFPDNI 718
```


C:Genetics:
A:Gene: yaer
C:Superfamily: protective surface antigen D-15

Query Match 3.8%; Score 178; DB 2; Length 810;

Best Local Similarity 17.7%; Pred. No. 0.03;

Matches 161; Conservative 133; Mismatches 311; Indels 304; Gaps 37;

```
OY 139 IPEYGEQEPNSEVVVPTLEPEKGLIKRLYA-----RLFNDS-----VNVKVRLLKA 185
DB 46 MVRRTGDTVNDNDI-----SNTRALFATGNEFEDRVLRDGTLLVQVERPTIAS 96
OY 186 KFYSSQSGETS-----AIGSSHQKTEPYANIKAALEDITQESAMDLNGSIPRL 234
DB 97 ITFSGNKSVKDMKLKONLEASGVAVGESLDR-TIADIEKLEDF-----140
OY 235 RQTAIVAAVAGVYDIDLSIINSGEVENIITHDLGPV---YIDRAVEVREGAD---288
DB 141 -----Y-----SVGKYSASVKAAYTPLPRNRVVDLKLVEQGVSAEIQ 179
OY 289 -----DKAFTT-----VADEVPL-LIGDFVHHGKETKKNLEMSAHEGYFDGRW 333
DB 180 INIVGNHAFNTDEBLISHPOLRDEVPMMNVGD---RKIQOKLAGDLETLNSYIDRG 235
OY 334 LDRSVVILLPONTADVSLIYDT-----GTQYRPEVVFETIDPKTNQLTTPDKLPVRE 388
DB 236 ARFNIDSTQVSLTPDKKGIYTVNITEGDQYKLSGEV-----SGNLGHSA 282
OY 389 LLEOLLTVNMGAVYVLAVALSNDL--IATRYFNNVNTIYEPPEEOIQNDQVSEQSS 446
DB 283 EIEOLTKIEPGLYNGTKYTKMEDIKLLGRY-----GYAVPRVQSP-----EIND 330
OY 447 SSRPEAOVDEESTLPRVLETFVELTDCILMDISPIEFSASNIODKNTLVAAKARHLYDM 506
DB 331 ADKTYKFLKVN-----VDAGNRYFYRKIRREGDTSKD-----AVLRREKOME 373
OY 507 DDRLVLAIHDDGVNRSILGRISDAVASAVARAILPSESENEVIDLPRTALANKRTPADY 566
DB 374 GAMLGSDLDVQKER--LNRLGFFETV-----DTDQGRVGSPPQVYV-----Y 416
OY 567 OSKKVPLVVFVASDKPRGQIGLWGSDDGTGLVTKFEHNLINRDGYQAGAEIRLSEDK 626
DB 417 KYKEKNTGSF-----NFGIGYGTESGVSPQAGVQODNMLGTGYAVG---INGTKN 463
OY 627 GVKLYATKPLSHPLNDQLRATLGYOQEVFGHSTNGFDSLSTRLEHISITLIONGMNRT 686
DB 464 DYQFYAEISVTNPTFYGVGSLGRLFTNDPQADADLSDTY-----NKS 508
OY 687 YSLRYRLDKLKTQAPPEWQDLPVDFVNGKPSQE--ALLAGYA-VHKTVADNLVNMGRY 743
DB 509 YG-----TDVTLGFPINEXNSLRAGIGYVHNISL-NMQPVAMW 546
OY 744 RORYSLEVGSSGLVSDAMKAT-----ARAGISGVYSF--G 776
DB 547 RYLYSM-GEHPSTSDQNSKFTDFTFNYGWTYNKLRGYEPTDGSRYNLGKTYIPGS 604
OY 777 DNAY-----GSNRAHOMTGGIOAGYIMSDNHNHYRLRFPAGGDSIRGA 823
DB 605 DNEYKYKTLDTATYVPIDDDHKMYVLGRTKMGYIGDLGKEKPEFYENFYAGSSSTVRQ 664
OY 824 HDLSLSPIS-----DKGY-----LTGGOVLAVGTAEY-----849
DB 665 SNTIGPKAVYVPHOASNVDPDYDECATQDQAKDLCKSDDAVGNAMAVASIEFTTPP 724
OY 850 -NYEFMDRLAVYGGDIGNAYDKG-----TNDTKIGAGYGVNRASPVGOVR 895
DB 725 ISDKYANSVRTSFEWDMGTVDITWMSQYSGPYDSDPSNIRMSAGIALQMSPLGLY 784
OY 896 VDVATGYKE 904
DB 785 FSYAQPFKK 793
```

RESULT 36

C70412

outer membrane protein - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999

C:Accession: C70412

R:Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;

V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MOID:98196666; PMID:9537320

A:Accession: C70412

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-778 <NOF>

A:Cross-references: GB:AE000733; NID:g2983720; PIDN:AAC07292.1; PID:g2983730; GB:AE00

A:Experimental source: strain VFS

C:Genetics:

A:Gene: omp

Query Match 3.8%; Score 177.5; DB 2; Length 778;

Best Local Similarity 19.9%; Pred. No. 0.03;

Matches 179; Conservative 122; Mismatches 377; Indels 221; Gaps 37;

```
OY 115 TTPLSLEELFAOESTENGINPNDIPE-----YQGEQPNSEVVVPTLEPEKGLI 165
DB 7 TLALFLGVANAALIEKTIIGNKYIPDELIRGLSVKSGSEFSLEIRKDIRKLYRTGLF 66
OY 166 KRLYARLFNDGVNKKVPRL-----KAKFYSSQSGETSAGSSHQKTEPYANIK 214
DB 67 TKVEAVEEDQNKKGPEVILLVEDLPYTIKEFEGNE-----FSDDDKLEYLGIET 118
OY 215 ALSDITQESAMDINGSLPRKOTALV-----AARANGYVIDLSIRNSIGEVDVITHL 269
DB 119 ELGEVEVEETIEGYTSSPAIEERLAWMKOLKILITFOEIE--SLRRRIEVEYKRRGF 176
OY 270 GEEVYIDYRAVEVREGGADKFTVADEVPLLIGDFVHHGKETKKNLEMSAHEGYF 329
DB 177 GTRV--EYKILPKK--GASLVE-----VIHGK-KIYTSDEIBGNKH--F 216
OY 330 DGRMLDRSVVILLPONTADVSLIYDTGTQYRPEVVFETIDPKTNQLTTPD-----381
DB 217 SDRLKGLMELKPN-----IFLRHPAPSEBTKLEDKYKIEEY 257
OY 382 -----KLPRKRLDQ-----LLTVNMGAVYVLAVALSNDL IATRYFNNVNTIY 429
DB 258 KNEGFLEVEVSYEVKKEGPARKYIVKKEGPRYKLEIKIEGNTLFA--YSELVDNLTK 315
OY 430 PER-----EOTON-----DQVSPQSSS---RREPQVDESTLEPIYEVLELT 470
DB 316 NERKGRYFRREVIEKTKNIREKYAIEIGPLNTSVEERVNVPKKEVSYLKIIE-----370
OY 471 DGLIMDISPIEFSASNI IODKNTLVAAKARHLYDMDDRVLAJNHDDGVNRSILGRISDA 530
DB 371 -GKPYVYKKIKIKIGNETHD--YVIREMK-----VOENLALAK--KGIRSS-----412
OY 531 VSAVARAILLPDSENEVIDLPRTALANKRTPADYQSKVP-----LYFVASDKPR 563
DB 413 -----KTRIMNLGYEDV-QIEVPPRDAMMDLLVIRIRFTG 449
OY 584 DQIGIGWGSDDGTGLRVTKEHNLINRDGYQAGAEIRLSEDKGVLYATKPLSHPLNDQ 643
DB 450 QFSVGLSTYEVTLGAGFIELRKGNFRTGDIAGISVYG-----SLYRNMAIS-----497
OY 644 LRATLGYOQEVFGHSTNGFDSLSTRLEHISIRSIIONGMNRTYS---LRYRLDKLKTQA 700
DB 498 -----YTRKMWLKKRPVDDLSAFDRIRIYDYTYVARTGFSVALSELSEYMRASIGT 551
OY 701 PPEWQDLPVDFVNGKPSQALLAGYAVHKTVAADNLVNMGRYRORYSLEVGSSGLVSDA 760
DB 552 GRVKYSDDIDPEASTYVKEQ-----AGRDRSRKIFFTLTRDTRDY---YLLPJKGSLFVFN 604
OY 761 NMAIRAG-----ISGVYSPGDMAVGSNRAHOMTGGIOAGYIMSDNHNHYRPLRKF 811
```

```

Db      605 SVAGGLGDEKFKFEVGEAKYFSDFYFDGIIILSLKG--EVGFEVGEYGGKKVPIDEBF 662
Oy      812 FAGGDOSIRGVAHDSLSPISDKYL--TGGOVLANGTAEVNYE-FMKDLRLAVFPDIGNA 868
Db      663 FVGGDFSIKRGYEVYAGTV-DPNTLDPIGAKKVTTSLELMVPFKKMLYFAGFPDYDYG 721
Oy      869 YDKGFTNDTKI----GAGVGRMASPVGOVRDVATVKE---EGNPDKLHFFIGTFP 919
Db      722 ADK--WSDFKLSNFRGGYGVIGVIRITTPAPIRIDMAFKTKIVPGDNTNSRIHFVLTGTF 778

RESULT 37
surface antigen VC2252 [Imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82099
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: B82099
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-803 <HEI>
A:Cross-references: GB:AE004297; GB:AE003852; NID:96656810; PIDN:AAF95396.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Gene(s): VC2252
A:Map position: 1
C:Superfamily: protective surface antigen D-15

Query Match      3.8%; Score 177.5; DB 2; Length 803;
Best Local Similarity 19.4%; Pred. No. 0.032;
Matches 174; Conservative 128; Mismatches 340; Indels 255; Gaps 40;

Oy      139 IPEYGEQPNSEVVPPTLEPEKPKGLIKRLYA-----RLFNDGVKKVPRLK----- 184
Db      46 MPKRVQDSVDSQVA-----NIIKALYSGNFEDVKYLRGNTLMVQVKRPTIAS 96
Oy      185 -----AKFYSSOSGSETSAI--GSSHOKTEPYANIKAALEDITJOSAMDLSIPRL 234
Db      97 VSESGKKAIEEQQLKLEASSIRVGEALDR--TLSNIEKGLDFY-----SVGKY 147
Oy      235 RQALYAARAVGYVDLDS--IIRNSIGEVDYIIHDGEPYIDYRAVEVREGADDKAFT 293
Db      148 NATVKAUVTPLPNNRADLKVFTEGVSAKIQIQINFGNQVFSDELLSRFLNLDVAMWN 207
Oy      294 TVADE---VPLLIGVFHNGKYETKKNLJENASAEHGYFPGRMIDRSVD---VILPNT 346
Db      208 FLADDDKKYQVLAGDLEALRTY-----YDRGTYLKFOVDSTOVAISPDKK 252
Oy      347 -ADVSLIYDTGYQVRFDEVVFTIDPKTNQLTJDPDKLPVKRELLLEQLLVNMGAEVNLQ 405
Db      253 GVAITPLNMEGEVYTSVKQF-----RGLMGKKAEMFTSLIFELGETNGS 299
Oy      406 AVRALSNDLIATRFMMVMTETVEFPEREQIQNDQVFEQSSSRTEPAQVDESTLEPYIE 465
Db      300 AVTRLEESV-----KKVLEGSGVAYPQ---VRIIPFEDER--QOVS 337
Oy      466 TVELTDGILMDISPIEFSSANLIQDKLNLVAAKARHLYPDPRDRLVALINHDGCVNRSLG 525
Db      338 VVHVEGKRVYRDITRFVGNKNSRDEY-----LRREKQWEGSWLSKDIETKTR--LN 390
Oy      526 RLSDAVSAAVARAILPDESENEVIDLPERALTANRKTPADYQSKVPLVYFVASDKPRDG 585
Db      391 RLGFETVEVQTVRVVPGSEQV-DL-----VYSVKEA-----NSG 424
Oy      586 QI--GLGMSDPTGTRLVYTFEHNLIINRGQYQACAEILRLSEDDKKGVKLVATKPL----- 636

```

```

Db      425 NVNFGVGTESGVSPFQVGLQDQNFILSGNRVGNVAMINDYQNLTLTEXRPMYMLDGV 484
Oy      637 --SHPLNDLRAIT-----LQYQOEVEGHS--TNGFDLSTRLEHISIIQNGSMNTTSL 689
Db      485 LGGKVFYNOFSEASEAGIVDYTNESYGTSLTWGFPFD-----ELNRF-----EFGI 529
Oy      690 RYRLDKLKTQAPPETWQDDLPVDFVNGKPSQEALLACVAVHKVYADMLVN----- 738
Db      530 GYTHKIKIGLTP-----YIQV-----ENFLAQAASINDSGNLLDQDPINLSWTR 575
Oy      739 -----PMRGYRQRYSLVSGSGSLVSDANMAIARAGISGVYSPGDMAVGSNRAHQMT- 789
Db      576 NMLNNSYFPPTAGNHQRAFYKMTVPG--SDAQYFKIQDYVRQYFPL-----TKKHEPFL 626
Oy      790 -----GGIOAGYIWSDNFHHV--PYRLRFPAAGDOSIRGAHDSLSP-----SDKG 835
Db      627 ILRGRLGNGNGYQDQKKNLPFEYENFVAGGFTSLRGSGSNSAGPKAVRYDYSGNNS 866
Oy      836 LT-----GGVLAVGTAEY-----NYEPMKDLRLAVE-----GDIG 866
Db      687 DTATDSDVGNMIALASVELIYPTPPASEARNQITTSIFYDMASVWDTPEFDRGKADYG 746
Oy      867 NAYDKGFTNDT--KIGAGVGRMASPVGOVRDVATVKE-EGNPDKLH-FFIGTFP 919
Db      747 NQYVYDSDPTMYRASSYVALQWSPMGPLVPSLAKPIKKYGGDDDEFFTFITIGTF 803

RESULT 38
JC4078
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)
C:Species: Haemophilus influenzae
A:Variety: type b
C:Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
R:Flack, F.S.; Loosmore, S.; Chong, P.; Thomas, W.R.
Gene 156, 97-99, 1995
A:Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus 1
A:Reference number: JC4078; MUID:95255676; PMID:7737523
A:Accession: JC4078
A:Molecule type: DNA
A:Residues: 1-797 <FLA>
A:Cross-references: GB:U13961; NID:9537447; PIDN:AAA85645.1; PID:9537448
A:Experimental source: type b
C:Superfamily: protective surface antigen D-15
C:Keywords: surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match      3.6%; Score 171.5; DB 2; Length 797;
Best Local Similarity 19.8%; Pred. No. 0.069;
Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

Oy      263 DVIHDLGEPYIDYRA-----VEVRGECA-DDKAFTTVADEVPLLIGVFNHKGYETK 315
Db      72 DVKAHOEGDVLVSVAKSIISDVKIKGNSVLPTEALQNQNDANGFKGVLDL-----IREK 127
Oy      316 KNLENSAEHGYFDGRDLRSVDYI--LPDNTADVSLIYDTGYRFDEVVFTIDPK 372
Db      128 LNEFAKSVKEHYASGRV-NATVEPIYVTLPPNNRAEILIQINEDDKAKLASLTF-----K 181
Oy      373 TNOLTJDPDKLPVKRELLLEQLLVNMGAY-----NLQAVRALNSNDLIATRFNMVNT 425
Db      182 GNE-SVSSSTIQEOMELQPSDWKMLGKKEFGAOFEKDLQSIK-----DYIINNGYA 232
Oy      426 EIVPEREQIQNDQVSEQSSSRTEPAQVD-----ESTLEPYIETVELTDG 472
Db      233 KAQITKTIDVOLNDEKTKVNTIDVNEGLOYDURSAARIIGNLGSABELEPLSLHLNDT 292
Oy      473 I-LMDISPIEFSSANLIQDKLNLVAAKARHLIYMPD--DRVLAINHDDGVNSIITL 524
Db      293 FRNSDIADVE---NAIKAKRGEGYSATVNSVPDEDDANKTLAITLVYDAGRRLTVRQ 348
Oy      525 -----GRISDAVSAAVARAILPDES--ENEVIDLP-----ALANRKTPAD----- 564

```

```

Db      349 LREGTVASADSTLRQEMROQECTWYNSQLVEIGKIRLDRGTGFEETVENLIDPIGNSDE 408
      565 ---VYOSKRVPLVYVAAADKPRDGOI--GLMGSDTGTRLVTKF-EHNLI----- 608
      409 VDVVYVK-----ERMTGSINFGICGTGESSISQASVKODNPLGTCAANSIAG 457
      609 NRDCYAGAEELRLSE---DKKGVKLYATKPLSHPLNDQLRATLGQOQVEGHS-TNGFDL 664
      458 TKNDYGTSVNLGYTEPEYFTKDGVSLSGNVFEFENYDNKSDTSSNKRTYTGSNVTLGEPV 517
      665 SFRF-----LEH---EISRIITONGGMRNT---YSLRYRLDKLKTQAPPETWODLPVDFV 713
      518 NENNSYYVGLGHTYKNKISNFALF---YNRMLYIOSMKFMKNGIKTN-----DEDFS 565
      714 NGKPSOELLAGVAHVKTVDNLVNPBRGRYRQRYSLSEVSSGLV---SDANMALIARAGI 769
      566 FGM-----NYSNLNRGIEFPKGVKAKASLGGRYTIIGSDNKKYKISADY 607
      770 SGVYSFGDNAYGNSRAH--QMTGGIOAGYIWSDFNHPVRLRFEFAGDQSIGRYAHDSL 827
      608 QCFEYPL-----DRDLWVVSAKASAGYANGFGNKRRLPFYQYTYTGGIGSLGFAVGS 660
      828 SP-----ISDKGLTGGQVLAAGTAE--VYEFMKD-----LRLAVF 862
      661 GPNATYAEYNGSGTGTEFKKISSD--VIGGNATATASAEILVPTPEVSDKSQNTVTSLSF 718
      863 GDIGNAY-----DKGFTNDTKIGAGVGYRMASPGVQAVDVATGV 902
      719 VDAASVWNTKMKSDKNGLESVDLAKRLPDYGSKRIRASTVCGVQWOSPISGLPVSTAKPI 778
      903 K--EESGNPIKLHFFIGTTPF 919
      779 KKYENDVDEOFOSIGSF 797

Db      779 KKYENDVDEOFOSIGSF 797

RESULT 39
D70304
hypothetical protein aq_050 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: D70304
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: D70304
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-853 <AQF>
A:Cross-references: GB:AE000671; NID:g2982793; PIDN:AAC06430.1; PID:g2982806; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: aq_050

Query Match      3.6%; Score 171; DB 2; Length 853;
Best Local Similarity 19.4%; Pred. No. 0.082;
Matches 179; Conservative 136; Mismatches 325; Indels 284; Gaps 47;

      169 YARLFNDGVNKPRLKAKFYSSGSGETSAIGSSHQKTEPYANIK-----AALEDITQ 221
      39 YAKIL-EMLKQPEIRREVYERKRKDLIIV-----ERYPIIKIKIKGNVFLNDEVD 90
      222 ESAMDLSIPRLQRTALVAARAVGYDDLSIIRNSIGEVYIHDGEPV-YID----- 276
      91 KNVIGIEGIPLEIDNPPTYEIRLEKRYKEIGFL-NADKRVNISLDDKNAIFYINIRG 149
      277 --YRAVEYRGEGADDAKFTTVADEVPLILGDFVHHGKYFTKKNLNLEASAEIYFGPGRML 334
      150 NLVLEFDIQFESVKKLQKKELKASGLVIGSIFDIDKVDADAEENLENFRKKGFFE----- 205
      335 DRSDVILIPDNTADVSLIYDNG-----TQYRDEVVFFTIIDPKTN-----QLTPT 378
```

```

Db      206 -----SEVYLIGKIKKELKRRFRALF-----PETDSFLKSLSIGLKNLVN 246
      379 DDDKLPVKRELLEQ-----LLTYNMGNAVNLQAVRALSNDLITRY-----ENMATE 426
      247 HP--LATLKAIGKGVGIPVYEEYEEYEL---KFLGNFFSDELYLSLEFDINTVGVD 301
      427 IVPEE--REQIOND-----QVSFE-----OSSSRTPAQV-----D 456
      302 ILVLEGKDKKEERYKRRKGFVDVAVEYLEBHRLLIKIKBEERKAKAYINGKRIELPYD 361
      457 ESTLEPVIE-----TWELTDGLMDISPIEESA-----SN 486
      362 EEKIQDLSIKREIQYLEKMGVYATAYELKKINKDKKEVNAVHKINKGRYIWSFKIESE 421
      487 LIQD--KLNLVAAKAKHLIDMPDRVLAIN-----HDDGVNRSI-LGRISDANVAS 534
      422 LFEDLNKRIISLKFPAKD-YETLDEIYKEIKRLREKGYFAKVFPTDIRQKVDALIMF 480
      535 ARA-ILPDE-----SENEVI-----DLPERTALANRKTPAVYQ 567
      481 YKAVVITGEERYEGDTLILYLEKTRLKEAEVLEKEKTFSEVERSW--NAIESEIFK 538
      539 SLRLIEDYVDRSKKVRHLAVFOEKKRGVIGLSAGFNT-----FE-----GFKLSAE 584
      568 SKKVPVYVVAADK-----PRDQIGLGWSDTGTRLVTKFEHNLIINRDYQAGAE 618
      619 LRLESEKKGVKLYATKRLSHPLNDQ-----RATLGQOQEVFGSTNGFDLSTLTLEH 673
      585 LSL-RNLIGLIGLINTFNFSISEKELYRVSPKDNFLRSRFFETSLKDYE---HDT 639
      674 SRSLIONGGWNRRYSLSRYRLDKLKTQAPPETWODLPVDFVNGKPSQ--EALLAGVAHVKT 732
      640 YELFTE---GFAFSLGIRLK-----PTTFGIGISNFEAKTTGERDKGT 681
      733 ADNL-----VNPMBGRORYSLEWSSGLVSDANMALIARAGISGVYSGDNAG 781
      682 YRKLGLSFSYKERPRFLVLFPGYGHRSKITELEKIK--KEFEKPCSRLKFSVG---YA 736
      782 SNAHQMTGGIOAGYIWSDFNHPVRLRFEFAGDQSIGRYAHDSL-SPIISDKCYLTGGQ 840
      737 YKKA-----PIFERPFLGGYKMKKGTYESIGSPL-----GGR 769
      841 VLAIVTAEVYEFPMKDLRLVVF---GDIGNAYDKGFNDNR-IGAGVGYRMASPGVQAV 896
      770 QMLYISPIRYLLANRNLDELITFLELKVENKFPSLYKMKKIDIGFSAGR--TPVGLIRG 827
      897 DVATGVKE-EGNPIKLHFFIGTTPF 919
      828 DIAPVLEDFVKRPSRLKFLYLSVEF 851

Db      828 DIAPVLEDFVKRPSRLKFLYLSVEF 851

RESULT 40
AD2445
hypothetical protein all5116 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AD2445
R:kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2445
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-588 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA076815.1; PID:g17134254; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all5116
```


GenCore version 5.1.4-P5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:15:24 ; Search time 28 Seconds

(Without alignments)
1361.312 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 4727
Sequence: 1 MSKPVLFANRSEFPVALAAY.....TGVEKGNPKLHFFIGTPEF 919

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	436.5	9.2	578	1 YTFM_HAEIN	P44038 haemophilus
2	433.5	9.2	577	1 YTFM_ECOLI	P39320 escherichia
3	180.5	3.8	793	1 D153_HAEIN	O32629 haemophilus
4	178	3.8	810	1 UP05_ECOLI	P39170 escherichia
5	171.5	3.6	797	1 D151_HAEIN	P46024 haemophilus
6	167.5	3.5	795	1 D152_HAEIN	P44935 haemophilus
7	163	3.4	1325	1 BCC3_ACEXY	O9w763 acetobacter
8	159.5	3.4	1319	1 BCC1_ACEXY	O9w763 acetobacter
9	159	3.4	475	1 YDF6_SCHPO	P19450 acetobacter
10	150.5	3.2	991	1 SC44_RICAP	O10478 schizosacch
11	149	3.2	1022	1 SC44_RICAP	O9a164 rickettsia
12	138.5	2.9	5430	1 ACP7_HUMAN	O9u049 rickettsia
13	137.5	2.9	495	1 G6PD_STRPN	O9u043 homo sapien
14	137	2.9	900	1 TP38_ARATH	O54337 streptococc
15	136	2.9	6486	1 TYCC_BACBR	O49160 arbidopsin
16	134.5	2.8	1279	1 TYCC_BACBR	O30409 b tyrocidin
17	134	2.8	1036	1 BCSC_PSEFL	P58937 pseudomonas
18	133.5	2.8	1296	1 P200_MYCPN	P75211 mycoplasma
19	133.5	2.8	4349	1 ASAL_ENTFA	P17953 enterococcu
20	133	2.8	3726	1 FAF72_HUMAN	O9u9q8 homo sapien
21	132.5	2.8	856	1 TRX_DROME	P20605 dirosophila
22	132.5	2.8	896	1 CLPB_HAEIN	P44403 haemophilus
23	132	2.8	612	1 YD86_SCHPO	O10411 schizosacch
24	131.5	2.8	896	1 HS75_KLUMA	O14170 kluyveromyc
25	131	2.8	1434	1 RA50_THEAC	O9h1r8 thermoplas
26	130	2.8	981	1 PROB_UREPA	O9p9v6 ureaplasma
27	130	2.8	2738	1 SC44_RICPE	O9a137 rickettsia
28	129.5	2.7	1616	1 PCGV_RAT	O9a824 rattus norv
29	129	2.7	1115	1 SLAP_BACCI	P53126 neisseria m
30	129	2.7	1185	1 FRPA_NEIMC	P23226 dirosophila
31	128.5	2.7	880	1 MAPX_DROME	O34996 bacillus su
32	128.5	2.7	999	1 DP01_BACSU	P32926 homo sapien
33	128.5	2.7	2869	1 DSG3_HUMAN	O00798 plasmidium
				1 RBP1_PLAVB	

34	126.5	2.7	1302	1 FRPA_NEIMB	O9K0x9 neisseria m
35	126.5	2.7	1326	1 BCC2_ACEXY	O92861 acetobacter
36	126.5	2.7	1829	1 PRPC_NEIMB	O91y45 homo sapien
37	126.5	2.7	2136	1 RPI_HUMAN	P56715 neisseria m
38	126.5	2.7	2198	1 YLJ2_CABEL	P34367 caenorhabdi
39	126.5	2.7	3396	1 PCGV_HUMAN	P13611 homo sapien
40	125.5	2.7	1616	1 P200_MYCCE	O49429 mycoplasma
41	125.5	2.7	928	1 HX42_HAEIN	P45354 haemophilus
42	125	2.6	963	1 TBES_THETH	O06458 thermus the
43	125	2.6	1433	1 REST_CHICK	O42184 gallus gall
44	124.5	2.6	3433	1 UTR0_HUMAN	P46939 homo sapien
45	124.5	2.6	3712	1 LMA_DROME	O00174 dirosophila

ALIGNMENTS

```

RESULT 1
YTFM_HAEIN
ID YTFM_HAEIN STANDARD; PRT: 578 AA.
AC P44038;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein HI0698 precursor.
GN HI0698.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxId=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Geocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT rd."
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RT Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae."
RL Electrophoresis 21:411-429(2000).
CC -I- SIMILARITY: STRONG, TO E.COLI YTFM.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC EMBL: U32752; AAC2357.1; -
CC TIGR: H10698; -
CC Interpro: IPR000184; Bac_surfAg_D15.
CC Pfam: PF01103; Bac_surface_Ag; I.
CC Signal: Complete proteome.
CC FT SIGNAL 1 22 POTENTIAL.
CC CHAIN 23 578 PROTEIN HI0698.
CC SEQUENCE 578 AA; 65812 MW; 34F9AC189C505876 CRC64;
Query Match 9.2%, Score 436.5; DB 1; Length 578;

```



```

Db      263 -----R-IENTI-----268
Oy      567 OSKVPLVYEWASDKPRDGOIGLWMSDPTGLRTVYTKFHEHNLINRDGYOAAEILSEDK 626
Db      269 -----ETGVGSTDVGRVATKMKPKMMNSYGLSTTSISAPBQ 309
Oy      627 GVKL-YATKPLSHPLNDLRATLTGLOQEVFEGHSTNGFDSLTRLEHISITSIONGMNR 665
Db      310 TLDPSYKMPLLKNPLEQYYLVGGGFKRT---DLNDTESDTPL-VASRYMDLSSQWR 363
Oy      686 TYSLEYRLDKLTQAPPTWQDLPVDFNCKRPOSEALL-AGVAHVHTVADNLVNPGRY 743
Db      364 AINLWMSD-----HFQIGTWTMTLFPYGVAMISKTRSGGLMPTWGD 407
Oy      744 RQRYSLVEYSGSLVSDANMAIARAGISGVYFCGNAYAGSNRAHOMTGGIOAGYIWSDNFN 803
Db      408 SQRYSIDVSNAMGSDVDVSVFOQANWIRLXYR-----HNFYTRGLTGMETGDFD 460
Oy      804 HVPYTRLFREFAGDOSIRGITYAHDSLSPISDKGYLTGGVYLVAGTAEVYEFMKDLRLAVFG 863
Db      461 KVPDLRFPFAGDGRSIRKYKYSIAPKYXANGDLGASKLITGSLEYOYNTGKMWGAFEV 520
Oy      864 DIGNAYDKGFTNDRIKIGVGVWRASPVGOVURVAVGV--KEEGNRIKIKHFFIG 916
Db      521 DSGEAVSDIRRSDFRTGTGVGVWRKESPEVGRPKIDFAVPVADKDEHG---DQFV 572

```

RESULT 3	ID	D153_HAE1N	STANDARD:	PRT:	793 AA.
AC	032629;				
DT	16-OCT-2001	(Rel. 40, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Protective surface antigen D15 precursor (80 kDa D15 antigen) (D-15 Ag) (Outer membrane protein D15).				
OS	Haemophilus influenzae.				
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;				
CC	Haemophilus.				
OX	NCBI_TaxID=727;				
RN	[1]				
RP	SEQUENCE FROM N.P.A..				
RC	SPRAIN-PAK 12085;				
RX	MEDLINE-97427952; PubMed-9284140;				
RA	Loomrose S.M., Yang Y.P., Coleman D.C., Shortreed J.M., Englund D.M., Klein M.H.;				
RT	"Outer membrane protein D15 is conserved among Haemophilus influenzae species and may represent a universal protective antigen against invasive disease."				
RT	Infect. Immun. 65:3701-3707(1997).				
RL	-1 SUBCELLULAR LOCATION: Outer membrane.				
CC	-1 SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				

Query Match 3.8%; Score 180.5; DB 1; Length 793;
Best Local Similarity 20.1%; Pred. No. 0.015;
Matches 159; Conservative 121; Mismatches 306; Indels 207; Gaps 39;

```

QY 263 DVIHDLGEPYIDRA-----VEVREGA-DDKAPTVADVPLLIGVFHHGKETK 315
Db 72 DVKAHQEGGVLVSVVAASIIIDVAKIKGNSVIPLEAKKUMLDANGFVGVDL----TREK 127
QY 316 KMLINASAEHGYFPGRWILDSRVDI---LDPNTADVSLIYDTGTYRPFBEVFEFTDPK 372
Db 128 LMFEAKSVKEHVASGRV-NATVEIVYVTLFNNRAEIIQIINEDKKLASLTF-----K 181
QY 373 TNOLTTPDPLKPVKRELEQLLTYNMGEAY-----NLQAVRALSDLIATRYFNMVNT 425
Db 182 GNE-SVSSSTLOEQMELQPDMSMKKLMGKRFEGAOPEKDIAIR-----DYLNGYA 232
QY 426 ELVFFERQIQNDQYSEFQSSSSRTPEQVQ-----ESTLEPIEVLELTDG 472
Db 233 KAQITKTDVQLNDEKTVNTVITVNEGLOYDLRSARIITGNIGNSAELEPLLSALHNDT 292
QY 473 I-LMDISPIEFSASNLIQDKLNLVAAKARHLHYDMPD---DRVAIHHDDGVNRSIL---524
Db 293 FRRSDIAYE-----NAIKAKIGEGYGTITVYNSVPDDPDANKTLATITFVVDAGRRLTVRQ 348
QY 525 ----GRISDAVASARAILPDES---ENEVIDP---ERT---ALANKRTPAD-----564
Db 349 LRFEBNTVASDSTLROEHRQOEGTWYNSQLELICKIRLDRTFGEFEYENRIDIPIGNSNDE 408
QY 565 ----VYQSKKVPPLYEVASDKPRDQOI--GIGWGSDBTRRLVTKR-ENHL-----608
Db 409 VDVYKVKV-----ERNKTSINEGIGYGESGISTOTSIKONPLGTGAANSIAG 457
QY 609 NRDGYOAGAEIRLSE---DKKGVKLATKPLSHPNLQDLRATLIGQOEVGHS--TNGFDL 664
Db 458 TKNDYGTSVNLGYTEPPTKGYSIGGNIPEYNDNSKSDPSNSYKRTTYGSMVTLGFPV 517
QY 665 SIRT-----LEH---EISRSIIQNGCMNRT---YSLKRLDKLTKQAPPETMODLPYDVY 713
Db 518 NENNSYVYGCHTYNKNKISNFALE---YNRNLXYIOSMKFEKGKIGXTN-----DFDFS 565
QY 714 NCKPQOEALLAGVAVHKTVAIDLNVNPMHGRYQORVSLKVGSGVL---SDANMAIRAGI 769
Db 566 FGM-----NYSNLNRGRTPYPTKGVASIAGGRVITPGSDPNKTKYKLSADY 607
QY 770 SGVYSFGDNAYGNSRAHQ--MTGISIQAGYIWSDNFNHPVYRLRPFAGGDOSIRGYAHSL 827
Db 608 QGFYPL-----DRDRHWYVSAKASAGYANGFGNKRILPFGYQTATGIGISLKGPAVGI 660
QY 828 SP-----ISDKGYLT-----GGQVLYAGIAE--YTFEPMKD---LPLAFGDIQNA 868
Db 661 GENAIYAEHNGJTEKNISSDIVIGENATITJASAEILVPTPEVSOXSONTVPTSLFVAAASV 720
QY 869 Y-----DKGFTNDTKIGAGVGRMASPVQGVVDAVTAQVK--EEG 906
Db 721 WNTKMSKDKNLESKVLKDLDPYCKSSRIASISYGVGQWQSPICGPLVFSYAKRIKYEEND 780
QY 907 NPKILHFTITPFP 919
Db 781 DVEQFOFISGGSF 793

```

RESULT 4	UP05_ECOLI	STANDARD:	PRT;	810 AA.
ID	UP05_ECOLI			
AC	P39170; P39181; P77465;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Unknown protein from 2D-page spots M62/M63/03/09/T35 precursor.			
CN	YAPT OR B0177 OR Z0188 OR ECS0179.			
OS	Escherichia coli, and			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.			
CC	Escherichia			
OX	NCBI_Taxid=562, 83334;			
NP	[1]			
RP	SEQUENCE FROM N.A.			


```

DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria: Proteobacteria: gamma subdivision: Pasteurellaceae:
OC Haemophilus
NCBI_TaxID=727;
RN
  [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype B;
RX MEDLINE=95255676; PubMed=7737523;
RA Plack F.S., Loosmore S., Chong P., Thomas M.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of
  Haemophilus influenzae."
RL gene 156:97-99(1995).
RN
  [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Minn, and Eagan / Serotype B;
RX MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
  Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
  species and may represent a universal protective antigen against
  invasive disease."
RL Infect. Immun. 65:3701-3707(1997).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC
  -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC
DR EMBL: U13961; AAA85645.1; -
DR EMBL: U60832; AAB61974.1; -
DR EMBL: U60833; AAB61976.1; -
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL
FT CHAIN
FT POTENTIAL.
FT PROTECTIVE SURFACE ANTIGEN D15.
SO SEQUENCE 797 AA; 87675 MW; 2F93DE538696AF1B CRC64;

Query Match          3 6%: Score 171.5; DB 1; Length 797;
Best Local Similarity 19.8%; Pred. No. 0.046;
Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

QY 263 DVIHDLGEPYIDYRA-----VEVRGEGA-DKRAFTTVADEVPLLDGVFHHKYEYTK 315
DB 72 DVKAHQEGDVLVSVVAKSIIDVKTKGNSVPTETLAKONLDANGKVGVDL-----IREK 127
QY 316 KNLINASAGHEGYFDGRWLDRSDVVI--LPDNTADVSLIYTGQYRFDEVVEFTIDPK 372
DB 128 LNEFAKSYKEHAHSAVGRY-NATVEPIVNTLPNNRAEILQINEDDQAKLASLTF-----K 181
QY 373 TNLQTLTDDDKLPVKRELLEQLTLTVKGEVY-----NIQAVRAISNDLIATRYEMVNT 425
DB 182 GNE-SVSSSTTLOEQWELDPDSSWMKLMGNKFEGAOFEKDLQSTR-----DYLLNGYA 232
QY 426 EIVPEREOIONDOVSFEQSSSRTEPAQVD-----ESTLEPYIEVELTNG 472
DB 233 KAQITKTVOGLNDEKTKYNTITDVNEGQYDLRSARITGNLGMSELEPLLSALHLNDT 292
QY 473 I-LNDISTIEFSASNLIDDKLNLVAKARHLYDMPD---DRVLAINHDDGVNRSIL-- 524
DB 293 FRRSDIADVE-----NAIKAKIGERGYSATVNSVPDPDANKTLAITLVVDAGRLLTVNQ 348
QY 525 -----GRISDAVASAVARALLPDES-----ENEVIDLP-----ALANRKTTPAD----- 564
DB 349 LRFQGNVYSAUSTLQKQEMRQEGTWYNSQVLELGRILDRGFEFTVENRIDPINGSNDE 408

```

```

QY 565 ---VYQSKVPLVYFVVASDKRRDGOI--GLWGSQDTGRLTYTKF-EHNLII----- 608
DB 409 VDVIYKVA-----ERNKTSINFGIYGTGESGISYQASVQKDMFLTGAANVTAG 457
QY 609 NRDCYQAGAEPLRSE---DKGVKLYATKPLSHPLNDQLRMTGYQGVFGHS--TNGDGL 664
DB 458 TKNDYGTSVNLGYTEPYPTTKGVSLGKGVNFEENTDNKSDTSSNKRTTYGSNVTLGFPV 517
QY 665 STRT-----LEH-----EISRSIIQNGWNR-----YSLRRLDKLTQAPPETMODLPVQFV 713
DB 518 NENNSYVYGGLHTYNTKISNFALF---YNRNLXIQSMKFKKGIGITN-----DPEFS 565
QY 714 NGRPSQELLAGVAVHKTVDNLNVPARGYQORISLEKSSGLV---SDANMALIARAGI 769
DB 566 FGW-----NYSNLNNGYPTPTGKASLGGRYTIPGSDKKYKLSADV 607
QY 770 SGVYSGFDNAYGSNRH--OWTGGIOAGYISDNPNHVPRLRFPAGDOSIRGVANDSL 827
DB 608 QGFYPL-----DRDLHWYSASAKASAGYANGFGNKRRLPYQTYAGIGSLRGFAVGS 660
QY 828 SP-----ISDKGYLTGGQVLAAGTAE--YNYEFMKD-----LRLAVF 862
DB 661 GPNAIYAEYNGSGTGTFKKISSD--VIGGNAITFASAEILVPTTPYSDKSQNTVRTSLF 718
QY 863 GDIGNAY-----DKGFTNDTKIGAGYGVWASPVQGVRVAVATGV 902
DB 719 VDASAVNTRKWSKDNKGLIESDYLKRLPDYKSSRIASTGVGFQOMSPIGFLVESYAKPI 778
QY 903 K--EGNPKILHPICGPF 919
DB 779 KRIENDVEQFQSTIGSF 797

```

```

RESULT 6
D152_HAEIN
ID D152_HAEIN STANDARD; PRT; 795 AA.
AC P44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
GN HI0917.
OS Haemophilus influenzae.
OC Bacteria: Proteobacteria: gamma subdivision: Pasteurellaceae:
OC Haemophilus.
OX NCBI_TaxID=727;
RN
  [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
  Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
  McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
  Scott J.D., Shetty R., Liu L.-I., Glodek A., Kelley J.M.,
  Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
  Rutterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
  Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
  RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
  RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
  Rd."
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC
  -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/

```



```

Db 320 QINSHADSLGGLVSMRQDAEARRRYQEQMAADPKTADMRPALAMEISGDYAAV 379
QY 235 ROTALVA-----ARAVGY---DIDLIRNSICEVDYIHDLEGPYIDY 277
Db 380 RQ-LIAHQYTEAKORLSLARQPGCFGATIMLADLOFTQIDASEQ-----EX 429
QY 278 RAVEVR-----GEGADDKAFTTVADVEPLL--IGDVFHHKKYETKNLIE---NA 322
Db 430 RSLADPNNOLAMGLIARDMAQGNATAEAROLSRVGP-----QYATEGEIEVGLMA 484
QY 323 SAHEGFDRMLDRSDVILPONTADVSLIYDGTQYRFEDEVFFITDPTNOLITDPDK 382
Db 485 AASH-----TSOSARVAILREMTQAPDPWVRIWL--ANALQOQGDV 536
QY 383 LPVKRELLBOLLTVNNGEAVNLQA---VRALSNDLIATRYFMVNTVEIYF-----ER 432
Db 527 AENGGRV--QPLIANVTVQDRQAGILYTYGAGNDAATRLLSGLSPEDYSPAIRSIAE 584
QY 433 EGIQNDQVFEQSSSRTEPAQVDESTLEPVETVELFDGILMDISPIEFASNLIDOKL 492
Db 585 MOIKEDLAS---RLSWVPNPVPLIREALAP---PPTGARGVAVADLFRQGDMLIHARM 637
QY 493 NLVAAKARILYMPDDR-----LAINDDGYNRSILGRISD-AVSAAVARALLPDESE- 544
Db 638 ALRIASTRTIDSPDQRLAYATEYKMSNPVAAAR-LLAPLGGSGSGAGCNALLPEQOQT 696
QY 545 ----NEVIDLPERTALANKRTPADVQSKVPPLYVAVASDPKPDQIGLGWGSDDTGRLL- 599
Db 697 LQQLRMGIIVASDILLNQGDDQAQAVDLHAPALARADPEATSPRLALARLYNGEGKSKAL 756
QY 600 ---VTKFEHNLINRQGYQAGAEIRLSEDKKGVLYATKPLSHPLDOLRATLGY----OO 652
Db 757 DIDLAVLRHNPDDLARQAAVOAAVNSGRSLATHLAMGVQESPMADARWLGMVAADQA 816
QY 653 EYFGHST-----NGPDLSTRLEHEHSRTIQGWNRTYSLRYRDKLKTQAPPET--- 704
Db 817 DGHGHTTADLRAYDLRLQOVEGSRAS---GPAATEE-----DAL---APPSNPP 863
QY 705 -----MODLPVDFVNGKPSQEA-----LAGVAVH-KTVADNL---VNPMSGYRQ 745
Db 864 RHNGYGRQTELPARTGGSYSMEATSPLEAADOMLSSISOINTLRKRLAPSIDGGLGFRS 923
QY 746 RLSLEVGSGGLVSDANMAIA-----RAGISGYVSFGDNAYGNSRAHQMTGGIOAGYIWD 800
Db 924 R-SGEHG-MGRLTENIPIVGRPLQAGES-----SLFESTPTMIWGC 965
QY 801 NFN-----HVPYRLRPFAGDQDSIRGYAH--DLSPLISDKYL-----TGGVLAVGTAE 848
Db 966 DLNAGSYDVP---RY--GTNATEAYNOYVNSLSQNNSSSSSLRTQOIOGGQ----- 1012
QY 849 YNFERMKDLRLAVFGDIGNAYDKGFTNDKIGAGVVRWASPYGVAVRATGVYK 903
Db 1013 -----GEAGFAPDYQFSN-SWVRADVG---ASPLIGPITTVLGVGE 1049

```

```

RESULT 8
BCCL_ACEXY
ID BCCL_ACEXY STANDARD: PRT: 1319 AA.
AC P19450:
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose synthase operon protein C precursor.
GN BCSC.
OS Acetobacter xylinus.
OC Bacteria: Proteobacteria; alpha subdivision: Acetobacteraceae;
OC Glucanacetobacter.
OX NCBI_TaxID=28448;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1306-3;
RX MEDLINE=91045951; PubMed=2146681;

```

```

RA Wong H.C., Fear A.L., Calhoun R.D., Eichinger G.H., Mayer R.,
RA Anilkam D., Benzinan M., Gelland D.H., Meade J.H., Emerick A.W.,
RA Bruner R., Ben-Bassat A., Tal R.;
RT "Genetic organization of the cellulose synthase operon in Acetobacter
RT xylinum";
RL Proc. Natl. Acad. Sci. U.S.A. 87:8130-8134(1990).
CC -!- FUNCTION: Required for maximal bacterial cellulose synthesis. It
CC may be involved in the formation of a membrane complex for
CC extrusion of the cellulose product.
CC -!- PATHWAY: Bacterial cellulose biosynthesis.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: BELONGS TO THE ASCC/BCSC FAMILY.
CC -!- SIMILARITY: CONTAINS 9 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37202; AAA21886.1; -.
DR PTR: C43735; C43735.
DR InterPro: IPR003921; Cellsynth.C.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR: 3.
DR PRINTS: PR01441; CELLSENTHASEC.
KW Cellulose biosynthesis; Repeat; TPR repeat; Membrane; Outer membrane;
KW Signal.
FT SIGNAL 1 46
FT CHAIN 47 1319 POTENTIAL.
FT REPEAT 49 82 CELLULOSE SYNTHASE OPERON PROTEIN C.
FT REPEAT 84 116 TPR 1.
FT REPEAT 291 324 TPR 2.
FT REPEAT 325 338 TPR 3.
FT REPEAT 405 438 TPR 4.
FT REPEAT 557 590 TPR 5.
FT REPEAT 701 734 TPR 6.
FT REPEAT 736 768 TPR 7.
FT REPEAT 1059 1094 TPR 8.
FT REPEAT 1319 1414 TPR 9.
SQ SEQUENCE 1319 AA; 141431 MW; 48C6BBA6C95A7F7 CRC64;

```

```

Query Match 3.4%; Score 159.5; DB 1; Length 1319;
Best local Similarity 19.2%; Pred. No. 0.45;
Matches 201; Conservative 131; Mismatches 360; Indels 357; Gaps 51;
QY 43 PAHDTAI-----NQAKAGNPVLLT-PRQIQARL---NAAGLNARQSQAL- 84
Db 164 PPHSLAVEYQYQWAGVPAGMDQARAGIAGVAVSNPDQYRAQLAFQAQLTYNTSTBMGLT 223
QY 85 ----DVNFPDQSTI-SRIGSQSPPLGID-MSVIEETTPLSLELFAQESTEMGINPDYT 139
Db 224 RLKDLQSFQARVEAAMAAQSYRQTLSQLPVNPEQPLMEQMLSA-----HPND-- 273
QY 140 PEYQGEQPNSEVVPPLEPEKPGILIKRLYARLFNDGVNKPVRLLKAKFYQSQSGESAT 199
Db 274 ----TALREHMLHPGPGPRDKAGLARQ-----AGYQLAAGRLAAA 310
QY 200 GSSHQKTEPYANIKAALEDITQESAMDNLGSIPLRQTLAARAVGYDIDLSIIRNSI 259
Db 311 EQSFQ-----SALQINS-----HDAD-----SL 328
QY 260 GEVDYIYIHDLEGPYIIDYRAVEVRGEGADKATTVADVEPLLIGDVFHHKKYETKNLI 319
Db 329 GGGGLVSMRQGDRA--EARYFEEMAADPK-----TADRMRPALAGAVSGEYASVROLI 382
QY 320 ENASAEHGYFDGRW-----LDRSYDVY----- 341
Db 383 ----AAHQYTEAKQOLATLARQPGQYTGATIMLADLQRSIGQIAAADQERYGLISREPNN 438
QY 342 -----LPDNTADV-SLIYDTGTQY--RFDEVVFFITDPTNOLITDPDKLRYK 386

```



```

Db 439 QALMLGLARVDMAGNTAEAROLLRSVGPQVAGVEIENSGLMAAASQTSDSAKVSL 498
QY 387 RELLEQL-----LTVMNGEAYNLQ-----AVRAL-----SND 413
Db 499 REAMAQAPRPPWVRININLALMOQOGDVAEAGRWQPIIANPVTAODRQAGILTYGSSND 558
QY 414 LIATRYNMYNTEIVPEREQIQNDVSEFGSSSRTEPAQVDESTLEPIETVELDGI 473
Db 559 AMTROLLAGSPADYSPAIRSIA-EEMEIKODLASRLSMVSNPVLIREALTQDPPTGAR 617
QY 474 LMDISPIEFASNLIDQKLNIVAKARHLMDPDRV-----LAINDGVNSIIGRI 527
Db 618 GVAADVLFPROGQDVHARMALRIASTRTIDLSPOQRSLVATEYMKISPVAAAR-LIAPL 676
QY 528 SDAY-SAVARAILPDESE-----NEVIDLPERTALANKRTPADYQSKKVPPLYVFASDK 581
Db 677 GDTGSGATGSAALLPEQVOTLQQLRMGISTVAGSDLLNRGDAQVADHLARLQADPEATS 736
QY 582 PR-----DGOIGLGMGSDTGTRLVTKFEHNLINRGYQAGAEIRLSEDKKV--KL 630
Db 737 PKLALARLYNGHGKPKALEID-----LAVLRHNPQDLDAHOAAVQAANSNDHNSLATRL 791
QY 631 YATKPLSHPLMDQLRATLGYQGEVFGSTNGFDLSTRTLEHISRSIIQNGMWRYSLR 690
Db 792 AMDCVQSPMDARMLMAVADQDGH-----GQRTIE-----DLRRANDLR 833
QY 691 YRL-----DKLKTQAPPET-----MODLPVDFYNGKPSQEA----- 721
Db 834 LQOEGTGRAASGAGAAQEDAL--APPSTNPFPRGCHQTELAGAPVYGSGYSMAASPD 890
QY 722 ----LAGVAVH-KTVADNL--VNPARGYQKRSLEYSGGLVSDANMAIA-----RAG 768
Db 891 TSDOMLSISGQITTLRENLAPSIDGLGFRSR-SGEGH-MGRLEANIPIVGRLLPADG 948
QY 769 ISGVYSGDNAVGSNRAHQMTGCIQAGYIMSDFN-----HVPRLRFPAGGDSIRGY- 822
Db 949 ASA-----LFTSTPTMINSGLNTGSYVYF--R--GTMMGVQAYN 987
QY 823 AHDSLSPIISKGLYTG--OVLAVGTAEXYNEFPWKDLRLAVFGDIGNAVDKGFTNDTKI 879
Db 988 QYDS-----YTNAGRDQORIAAGTAE-----ACGAPADVOF 1017
QY 880 G-----AGGVNRASPYGOVRVDATGVK 903
Db 1018 GNSWVRADVG--ASPIGFPITVNLGVE 1043

RESULT 9
YDF6_SCHPO
ID YDF6_SCHPO STANDARD: PRT: 475 AA.
AC 010478;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C17C9.06 In chromosome I.
GN SPAC17C9.06
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21648401; Pubmed=11859360;
RA Wood V.V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham S., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Godle A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

```

```

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellion J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Wolckaert G., Aert R., Robben J., Grymoprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Scheifer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Filtz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Jelaune V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerretti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RA "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -I- SIMILARITY: BELONGS TO THE UPF0140 (CGI-51) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z730399; CAA97352.1; -.
DR InterPro: IPR000184; Bac_surfng_D15.
DR Pfam: PF011103; Bac_surface_Ag; 1.
DR Hypothetical protein.
SQ SEQUENCE 475 AA; 51762 MW; 8FEC5B5F4B5DE15 CRC64;

Query Match 3.4%; Score 159; DB 1; Length 475;
Best local similarity 20.8%; Pred. No. 0.1; Mismatches 136; Gaps 25;
Matches 113; Conservative 83; Mismatches 210; Indels 136; Gaps 25;

QY 442 FEQSSSRTEPAQVDESTLEPIETVELDGI LMDISPIEFASNLIDKLNIVAKARH 501
Db 5 FESTSFSDIPAVNESKLS-AETFKSUSEILANSTLPVGISSI----- 49
QY 502 LYMDPDDVLAINH-DGVNSIIGRID-AVSARAILPDESENEYIDLPERTALNR 559
Db 50 ----RVTGAHHTRPSTFIRVKLCTCDTSKPAKRSRL--ETLNAIOETTGNIAMFN- 99
QY 560 KTPADVOSKKVPL-----YVFVASKDPRDGOIGLGMGSDTGTRLVTKFEH 605
Db 100 ----VYETANIKIDIRASSVSGDDDLDTIYQKRP--LYVETGIDVG-NVSGNVHA 150
QY 606 NLINRGYQAGAEIRLSEDKKGVKLYATKPLSH--PLNDQLRATLGYQGEVFGHS----- 658
Db 151 NVLARNVF-GGAELLISGVNYSYTRNRSTMSVVFEPVNAADPTRLRFN---GHSNLRDN 205
QY 659 ----TNGFPLSTR-----TLEHISRSIIQNGMWR-TYSLKRLRLDKLTKQAPE 703
Db 206 KSISSHLLITLGTITLSLOHDIWSEHLLSONLL---WROVTHLFEASPSVRLEAGDS 261
QY 704 TWODLPVDFVNGKPSQEA LLAGVAVHKTVDNLVNPARGYQKRSLEYSGSG-LVSDANM 762
Db 262 LKQSLSTYIT-----RPTRHIMPTFGQDVYRQTLBELAGFLPGDASF 305
QY 763 AIAKAGISGVYSGDNAVGSNRAHQMTGIGIA--GYIWSDFNHNVPYLRFPAGGDSIR 820
Db 306 LKSE-----FMGQKAAVALNSSRSVSLTSARIGALHSLNKKQVSLDPRFMLGSGSTLR 358
QY 821 GYANDLSLPSIDKXVILGGVLAAGTAY-----NTEFPKDLRLAVFGDIGNAVYDKGFTND 876
Db 359 GFSEDRIGPKDGRSLGTATVAFMSLSLFPKPVDAKSPRLDIFAAAG-----GLISNL 413
QY 877 TK-----IGAGVNRASPYGOVRVD-----VATGQKEGNGNLIKLFHFGT 917
Db 414 TSPNPGCYKSLSKPCISTGLGVYANPAPARFELNPLIATTEKDIGR-GLOFGAGI 472

```


OY 918 PF 919
DB 473 DF 474

RESULT 10
SCA4_RICAU STANDARD: PRT: 991 AA.
AC 09A564;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia australis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=787;
RM [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RL 'gene D' coding for an intracytoplasmic protein."
CC Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF187982; AAK31302.1; -
KW Antigen.
FT NON_TER 1
FT NON_TER 991
FT SEQUENCE 991 AA; 108518 MW; F2EDB94DDC09749A CRC64;

Query Match 3.2%; Score 150.5; DB 1; Length 991;
Best Local Similarity 19.0%; Pred. No. 0.91;
Matches 186; Conservative 136; Mismatches 335; Indels 321; Gaps 45;

OY 45 HDTAINOK-----AGNPVULPPEIOARLNAAGLNAPQSOAL-DVNFDDQS--PISR 97
DB 7 YESGIDEKKOEKAAALQPTLDADGFSFTPASSTOSTPSISTLSDTLSHDQTSDDPIRK 66
OY 98 -----IGEOSPPGLDMSVIEETTPLSLELFAQESTEM--GINPNDYIPRY 142
DB 67 AVRETIIOPOKDEIAEQ-----IIKDLAALVDRDLADOKRKEIEEKKNTLSAF 117
OY 143 OGEOPNSEVVPPTLEPKPGLIKRL-----YARLFNDGVNKKV----- 180
DB 118 FGNPANRELIDKAL--EKPELKKLEAIETAGLKNVFLTYIAANGYSGGFPPVOMENI 174
OY 181 --PLUKAKFYOSSSGESATGSSHQKTEPANTAALEDDITQESAMDLNGSIPRLROTA 238
DB 175 SASDLRAIVVNDAGDELCTINETTKTKPPT--VAKODQVQ--VOIN----- 218
OY 239 LVAARAVGYVDIDLSIINSIGEVDVIHDGEPEYIDYRAVEVGEG--ADDKA--FT 293
DB 219 -----SYRIDPPI--KDKADGSMH-----LSMALADATKPSKDAVYFT 259
OY 294 TVADEVPLLGDFVHCKYETKKNLIENASAEHGYFDGRWLDRSVDVLLPNTADVSLIY 353
DB 260 AHYEGR-----NGRPOLK-----EISSPKPLKFAE-----DGRPAVAIE 295
OY 354 DTGQYRFDEVEFTTIDPKTQTLTTPDKPLFYKRELLEQOLLTVNKG--AVNLQAVRAL 411
DB 296 HGGRY-----TLAVTRGKYKEMKKEVELHOGOSVDLSQTIARDLTVKQGRS 342

OY 412 NDILATRFNMVNTIEYPEREDIQNDVSFEQSSSRTERPAQVDESTLEPVLETLD 471
DB 343 QETL-----QPIITPMOEL-----KSIETPTTQVPPITPDSOPLD--TETAQMPQ 387
OY 472 GILMDISPIEFESA-----SNLIODKLNVLAAKARNLXDPDRVLAIHNDGVNSILGRI 527
DB 388 S--QOVNPNLNNATALSSGMDLLNVNA-----GLTKKEDGNQI----- 427
OY 528 SDAVSAVARAILPDESENE--VIDLPETALANRKP-----ADV 565
DB 428 -DLINAAATGAILNNEKEQANIIVLTENTVNNALFPDTKVAGVNAVLENIKNNQNTPLD 486
OY 566 YOSKAPLYVFNVSOK--PRDQ-----IGLWGSDDTGRV-----TKRENH 607
DB 487 EKSKMLEATVALINSENLAPKOKQOLEKAVDVGMLKDDTSRAVALIDSTDTVIRKSNL 546
OY 608 INRDGYOAGAEILRSEDKGVKLYATKPLSHPLNDQLRATIGYQOEVFHSGTNGFDLSTR 667
DB 547 STRD-----KOTMLIAV-----GDKVASELSMA 570
OY 668 TLEHEISRTIIONGWNRTYS-----LRYRLDKLTQAPETWODLPVDFVNG--KPS 718
DB 571 E-KOOLLGSVYLKKGVEYQVLPPEQOOLMOONLDKITA--EQTNAKITEVGGILANPA 625
OY 719 QEALLAGVAVHKVADNLVNPNGYRGORYSLEVGSSGLVSDAMALARAGISG-----Y 772
DB 626 FNTIAKTEAIQNTTVKVLDSPIKAEIKGETLE-----STIKVVAESPLNODQKVDIV 677
OY 773 YSRGDNAYSNRAHOMTGGIOAGYIWSDNFNHVPYRLRFPFAGDOSTRGVANSLSPLSD 832
DB 678 KGAGE-AIASHKTMSTEKISA-----IESVEGVAESTIALED 715
OY 833 KGYLVGGVLAAGTAEVNYEFMKDLRLAVFGDIGNAYDKGFTNDTKIGAGVGVNMA SPV 892
DB 716 KRLMTKGLVDGIYEDKANPEMTKAVSGVDKSTARREDQALMDA-----AS 762
OY 893 QVRVDVAT-----GVKKE 905
DB 763 EVALDRETQNTFKGLKEQ 780

RESULT 11
SCA4_RICPR STANDARD: PRT: 1022 AA.
AC 09ZD49; 09ZD48; 09AJ36;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120).
GN SCA4 OR RP498/RP499.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RM [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Magrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
RM [2]
RP SEQUENCE OF 11-1016 FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RL 'gene D' coding for an intracytoplasmic protein."
CC Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

```

CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 234.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ235272; CAAL14951.1; ALT_FRAME.
DR EMBL: AJ235272; CAAL14950.1; ALT_FRAME.
DR EMBL: AF200340; AK31305.1; -.
KW Antigen; Complete proteome.
FT CONFLICT 11 15 EEPDL -> RPLGLV (IN REF. 2).
FT CONFLICT 365 365 H -> Y (IN REF. 2).
FT CONFLICT 413 413 MISSING (IN REF. 2).
FT CONFLICT 957 957 G -> R (IN REF. 2).
SQ SEQUENCE 1022 AA; 114410 MW; 03230E3A663A9622 CRC64;

Query Match 3.2%; Score 149; DB 1; Length 1022;
Best Local Similarity 20.7%; Pred. No. 1.2;
Matches 178; Conservative 131; Mismatches 303; Indels 248; Gaps 45;

QY 74 GLNAKRP-----SQALDVYVFDQSP--ISRIGE-----OSPP 104
DB 267 GPKGKQPKETSSPPKPKFVGTDGDAVAIEHGGEITLAVTRCKYEMKKEVALNHQGS 326
QY 105 LGDMSVIEETPTLSLELAEQESTEMGINPNDYIPETGQBPSEVYVPTLEPERGL 164
DB 327 VALSQTAEELTHVQGP---SHEHKKPIIIPNDEL-ESSIEGHNSQOVPRTT----- 374
QY 165 IKRLVYARLENDGVNKPRLAKFYOSSGSETSAIGSSHOKTEPYANIKALEDTIGESA 224
DB 375 -----TFNKSLQ--PKISQIHQLPQAQSSGI-----PNVLNAANALSTSMODLL 419
QY 225 MDLNGSIPR---LRQPLVAARAVG-----YYDIDLSIIRNSIGEVQY-- 265
DB 420 NNINSLTITKMODINKQSDLLKEAIAIILNNKSKDPAKQYNIIDLA--KNFSKKDITAD 477
QY 266 ----IHDLGEPYIDYRAVEVREG--ADDKAFTVADEVPPLIGVFHNGKYEKKNL 318
DB 478 AKVNVVNTLETIONDQNTLDIKKSKILEDTVAITLNSENIEL-----KQKQOI 526
QY 319 IENASAEHGFQDGRWLDKRSVDVILPNTADVSLIYDTGTQYRPEVVEFTIDPTNQLTT 378
DB 527 LE-----KVVYDI-----GLSTKDDISRVAVADSIMDTYI--KSNITANE 562
QY 379 DPKLPEVKRELEQLLTVNNGEAVNLQAVRALSNDLIATRFVNNVNTIEVPEREQION- 437
DB 563 DKRKIFI--TVFOQ---INSYFSSNAKKULLBSILKKT---AEQVLSPEQOOLMNO 612
QY 438 --DOVSFEQSSSRTEPAQVDESTLEPIETVELTQGI-----LMDISPIEFS-ASNLI 488
DB 613 MLDNITTEHTKKRPTIE--KVNNTILEPLSNTALKTNIQVMTSVNLD-SPVQIEMKSKLI 669
QY 489 QOKLNLVYAAKARLLYMPDORVLAINHDDGVNSIIGRISDA-----VSAVARAI 538
DB 670 QVVTKTVAESA--LVE-PKDKTEIVK--GIGKTIYVH--SDTSLPLHDKVYIMGSVAKGI 722
QY 539 LPDESENEVID-----LPERT-----ALANRKTPAOVYOSKVPV--VYFVA 578
DB 723 V--ESKNDLLDRLLIAGLVGDIYEAKGDMNAVYHAISSMANSINSSEKALKRSDVY 780
QY 579 SDRPRGQOI-GLGWSGDTGTRLYTKFEHNLINDGVQAGA-----ELRLSEDKKGVKL 630
DB 781 SEKVLDEKREIONDLRELKAONINESKLDHDIYKNKTODVANAALKVITVLDNDSQGRVSE 840
QY 631 YATKPLSHPLND-----QLRATLGYQ-----QEVFGHST-----NCFDLSTR 667
DB 841 EAPKVVSSLNDISKRTIEKINNLRAMLSDQGNLKTFEKKKDEATKVKVDELVKAFAFKSS 900

```

```

QY 668 TLEHE--ISRSIIQNGNNRTYSLRYLDKLTQAP---PEFMQDLPVDFVNGKPSQEALL 723
DB 901 TEEQONFIKSNLIDNKLTLSREIRLQIIDLNLKAQOKRAETENI-----SAKTEADVAVI 955
QY 724 AGVAVHKVAVANLVNPMGYRORSLEYSSGSLVSDAMMALIARAGISGVISGDNAYSN 783
DB 956 SGKSELKPIISD-----EPYIQKAMVVERDRVDIKDKIKMSALI-----NARDS- 1001
QY 784 RAHOMTGGIQAQYIMSDNFN 803
DB 1002 -----IQSENFN 1008

RESULT 12
ACF7_HUMAN STANDARD; PRT; 5430 AA.
ID ACF7_HUMAN
AC Q9UPN3; Q9UPK0; Q9UG9; Q9H540; Q75053;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Actin cross-linking family protein 7 (Macrophin) (Trabeculin-alpha)
GN (620 kDa actin-binding protein) (ABP620).
DE ACF7 OR ABP620 OR KIAA0465 OR KIAA1251.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP MEDLINE=20026884; PubMed=10559237;
RX Sun Y., Zhang J., Kraeft S.-K., Auclair D., Chang M.-S., Liu Y.,
RA Sutherland R., Saigla R., Griffin J.D., Ferland L.H., Chen L.B.;
RT "Molecular cloning and characterization of human trabeculin-alpha, a
RL giant protein defining a new family of actin-binding proteins.";
RN J. Biol. Chem. 274:33522-33530(1999).
[3]
RP SEQUENCE OF 868-2350 FROM N.A.
RX TISSUE=Brain; PubMed=10574462;
RC MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirosewa M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RN DNA Res. 6:337-345(1999).
[4]
RP SEQUENCE OF 1544-5057 FROM N.A.
RA Corby N.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 3734-5430 FROM N.A.
RX TISSUE=Brain; PubMed=9455484;
RC MEDLINE=98116662; PubMed=9455484;
RA Seki N., Ohira M., Nagase T., Ishikawa K.-I., Miyajima N.,
RA Nakajima D., Nomura N., Ohara O.;
RT "Characterization of cDNA clones in size-fractionated cDNA libraries
RT from human brain.";
RL DNA Res. 4:345-349(1997).
CC -1- FUNCTION: ACTIN-BINDING PROTEIN WHICH MAY PLAY A ROLE IN CROSS-
CC LINKING ACTIN TO OTHER CYTOSKELETAL PROTEINS. ALSO BINDS TO
CC MICROTUBULES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.
CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.

```

CC	-1	SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.	
CC	-1	SIMILARITY: CONTAINS 1 SH3 DOMAIN.	
CC	-1	SIMILARITY: CONTAINS 37 SPECTRIN REPEATS.	
CC		-----	
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-slb.ch/announce/or_send_an_email_to_license@isb-slb.ch).	
CC		-----	
DR	EMBL	AB029290; BAA83821.1; -	
DR	EMBL	AF141968; AAF06360.1; -	
DR	EMBL	AB030377; BAA6565.1; -	
DR	EMBL	AB107953; CAC15920.1; -	
DR	EMBL	AB007934; BAA22310.1; -	
DR	HSSP	Q01082; 1BKR.	
DR	Genew	HGNC:13664; MACF1.	
DR	InterPro	IPR001589; Actbind_actnin.	
DR	InterPro	IPR001715; Calponin-like.	
DR	InterPro	IPR002048; EF-hand.	
DR	InterPro	IPR003108; GAS2.	
DR	InterPro	IPR001452; SH3.	
DR	Pfam	PF000307; CH; 2.	
DR	Pfam	PF000307; ehand; 2.	
DR	Pfam	PF00435; spectrin; 36.	
DR	Pfam	PF02187; GAS2; 1.	
DR	ProDom	PD000012; EF-hand; 1.	
DR	SMART	SM00033; CH; 2.	
DR	SMART	SM00054; EFh; 2.	
DR	SMART	SM00243; GAS2; 1.	
DR	SMART	SM00150; SPEC; 35.	
DR	PROSITE	PS00019; ACTININ_1; 1.	
DR	PROSITE	PS00020; ACTININ_2; FALSE_NEG.	
DR	PROSITE	PS00021; CH; 2.	
DR	PROSITE	PS00018; EF_HAND; 2.	
DR	PROSITE	PS50002; SH3; FALSE_NEG.	
KW	Actin-binding;	Cytoskeleton; Calcium-binding; Repeat; SH3 domain.	
FT	DOMAIN	1 295	
FT	DOMAIN	78 181	CH 1.
FT	DOMAIN	194 295	CH 2.
FT	REPEAT	314 355	SPECTRIN 1.
FT	REPEAT	591 623	SPECTRIN 2.
FT	REPEAT	680 784	SPECTRIN 3.
FT	REPEAT	786 800	SPECTRIN 4.
FT	DOMAIN	871 923	SH3.
FT	REPEAT	1250 1272	SPECTRIN 5.
FT	REPEAT	1287 1342	SPECTRIN 6.
FT	REPEAT	1455 1534	SPECTRIN 7.
FT	REPEAT	1547 1659	SPECTRIN 8.
FT	REPEAT	1815 1891	SPECTRIN 9.
FT	REPEAT	1932 2042	SPECTRIN 10.
FT	REPEAT	2260 2280	SPECTRIN 11.
FT	REPEAT	2372 2395	SPECTRIN 12.
FT	REPEAT	2398 2507	SPECTRIN 13.
FT	REPEAT	2510 2618	SPECTRIN 14.
FT	REPEAT	2621 2728	SPECTRIN 15.
FT	REPEAT	2731 2838	SPECTRIN 16.
FT	REPEAT	2841 2945	SPECTRIN 17.
FT	REPEAT	2987 3024	SPECTRIN 18.
FT	REPEAT	3136 3163	SPECTRIN 19.
FT	REPEAT	3187 3274	SPECTRIN 20.
FT	REPEAT	3277 3383	SPECTRIN 21.
FT	REPEAT	3386 3492	SPECTRIN 22.
FT	REPEAT	3495 3601	SPECTRIN 23.
FT	REPEAT	3604 3673	SPECTRIN 24.
FT	REPEAT	3713 3819	SPECTRIN 25.
FT	REPEAT	3832 3927	SPECTRIN 26.
FT	REPEAT	3982 4043	SPECTRIN 27.
FT	REPEAT	4046 4152	SPECTRIN 28.
FT	REPEAT	4155 4262	SPECTRIN 29.

FT	REPEAT	4265	4371	SPECTRIN 30.
FT	REPEAT	4374	4481	SPECTRIN 31.
FT	REPEAT	4484	4590	SPECTRIN 32.
FT	REPEAT	4593	4700	SPECTRIN 33.
FT	REPEAT	4707	4808	SPECTRIN 34.
FT	REPEAT	4811	4917	SPECTRIN 35.
FT	REPEAT	4920	4985	SPECTRIN 36.
FT	REPEAT	5030	5054	SPECTRIN 37.
FT	CA_BIND	5096	5107	EF-HAND 1 (POTENTIAL).
FT	CA_BIND	5132	5143	EF-HAND 2 (POTENTIAL).
FT	DOMAIN	5276	5283	POLY-SER.
FT	DOMAIN	5355	5370	4 X 4 AA TANDEM REPEATS OF [GS]-S-R-[AR].
FT	DOMAIN	5370	5370	MSSSDSEETLSEKSCSEKSCSESRYSREKSGSLSPDGD.
FT	DOMAIN	5370	5370	TLPWNLPHLEOKRKSDSVLDPAERAIVRV -> MEPLYLM
FT	DOMAIN	5370	5370	ACIPGRDVGSLPLPPGFKCFCTASRAVAVI (IN REF.
FT	DOMAIN	5370	5370	2).
FT	DOMAIN	5370	5370	VAISSDEEGNIRFEVELLS -> GPSAPLKMAISDLCKNN
FT	DOMAIN	5370	5370	YCL (IN REF. 1).
FT	DOMAIN	5370	5370	A -> T (IN REF. 2 AND 3).
FT	DOMAIN	5370	5370	V -> A (IN REF. 1).
FT	DOMAIN	5370	5370	E -> D (IN REF. 1).
FT	DOMAIN	5370	5370	E -> K (IN REF. 2).
FT	DOMAIN	5370	5370	M -> V (IN REF. 2).
FT	DOMAIN	5370	5370	C -> Y (IN REF. 1).
FT	DOMAIN	5370	5370	MISSING (IN REF. 2).
FT	DOMAIN	5370	5370	SLIPSVG -> EYRFKI (IN REF. 3).
FT	DOMAIN	5370	5370	O -> R (IN REF. 2 AND 4).
FT	DOMAIN	5370	5370	S -> T (IN REF. 2).
FT	DOMAIN	5370	5370	MISSING (IN REF. 2).
SO	SEQUENCE	5430	AA: 620346	MM: 91AB7F7518B0440B CRC64:

Query Match	2.9%; Score 138.5; DB 1; Length 5430;
-------------	---------------------------------------

Matches 164; Conservative 137; Mismatches 288; Indels 327; Gaps 42;

Matches 164; Conservative 137; Mismatches 288; Indels 327; Gaps 42;

QY 14 PVALAAYLPMTSQALAQQ-----NNPANI-INHVPAHDTAINQAKAG-NPVL 60

Db 2873 PVQLES--SLRSKAMLNEVEKRSLLEILNSADILNSSEADEGIRDEKAGINQMD 2930

QY 61 LTPEQIQRUNA-----AGLNAKPQSQALDVNFDDQSPISRIGEQSPPL 10

```

Db 2931 AYTEIQAKTGSLEEMTORLREFQESFKNIEKKVEGAKHOLEIFD-----ALGSOA--- 29

```

QY 106 GLDMSVIEETTPLSLELFAOESTEMGINPN-DYIPEY-OG----- 14

```
Db 2982 -----CSNKLEKLRADQEEVIAALPEPOVDYI.RNFTGGLVEDAPDGSASQLTHOAE 30
```

0V 145 -----FQDPSFVVPPT.FPEKPGI-----TKRIYABIEN-----DCVNKKVPBI. 19

Db 3033 VAOOFEI-EVKOBVNSGCV--MENKIECTGOEHCVRBEMESOIADI DDEI DCMGATCB- 30

On 194 KAREVOSSCEMATA CCOUOMEDVAN - - - - - TPAED - - - - - TMOCHNDVNO 00

[illegible][illegible]

2	1	
2	2	—
3	3	—
4	4	—
5	5	—
6	6	—
7	7	—
8	8	—
9	9	—
10	10	—
11	11	—
12	12	—
13	13	—
14	14	—
15	15	—
16	16	—
17	17	—
18	18	—
19	19	—
20	20	—
21	21	—
22	22	—
23	23	—
24	24	—
25	25	—
26	26	—
27	27	—
28	28	—
29	29	—
30	30	—
31	31	—
32	32	—
33	33	—
34	34	—
35	35	—
36	36	—
37	37	—
38	38	—
39	39	—
40	40	—
41	41	—
42	42	—
43	43	—
44	44	—
45	45	—
46	46	—
47	47	—
48	48	—
49	49	—
50	50	—
51	51	—
52	52	—
53	53	—
54	54	—
55	55	—
56	56	—
57	57	—
58	58	—
59	59	—
60	60	—
61	61	—
62	62	—
63	63	—
64	64	—
65	65	—
66	66	—
67	67	—
68	68	—
69	69	—
70	70	—
71	71	—
72	72	—
73	73	—
74	74	—
75	75	—
76	76	—
77	77	—
78	78	—
79	79	—
80	80	—
81	81	—
82	82	—
83	83	—
84	84	—
85	85	—
86	86	—
87	87	—
88	88	—
89	89	—
90	90	—
91	91	—
92	92	—
93	93	—
94	94	—
95	95	—
96	96	—
97	97	—
98	98	—
99	99	—
100	100	—

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible]

20 2100 EVALUATION OF THE QUALITY OF LIFE OF THE STUDENTS OF THE 32

```

27      VADEVELLIGDVEFHNGNIE-----INNNLIEN--ASAEHGIFDGKW 33
28      : | | : | | | : | | : | | :
29      : | | : | | | : | | : | | :

```

DD 3236 WNIENNAQRIÄQLQEALHCGKFQDALEPLLSWLADIEELLIANQKPPSAEYKVKÄQI 33

334 LDRS-VDVILPBNIAVSLIYBIGIQREDEVEFTIDPKTNQLTTPDKLPVKRELE- 39

Db 3318 QEQKLLQRLDDPKATVDM LQAEGR I-----AQSAEL-ADREKITGQLESLES 33

QY 392 --QLTVNMGEAYNLOQVRAALS-----NDLIATRYENMNVNTEIVFPEREQIQ-- 43

```
Db 3366 RWTETLSKAAAROKOEDILVLAKOFHETAPISDFLSTYKKNLANSEFVGTOTAKIOO 3425
Oy 437 -----NDQVSFEQSSSSRTEPAQVDESTLEPIETV-----ELTDCI 473
Db 3426 IIRHKALEEDINHAIDVQAQVIGOSLSLTPA--EOGVISEKIDISQARYSETODRC 3483
Oy 474 LMDISPIEFSANL-----IIOKLNLAQAARHLYDMDDRLAINHMD 517
Db 3484 CKKALLDQALSNARLFGDEVEVLNWLAEVEDKLSVVK-----DFKOD-VLHRQHND 3537
Oy 518 --GVNRSILGRISDAVSAY--ARAILPDESENEVIDLPER--TALANRKPADVYOSKKVP 572
Db 3538 HLALNEIYVNRKKNVDAIKNGOALLKQTTGEBEVLIIQEKLDGIKRVRADITVTSKALR 3597
Oy 573 LYVFAASDKPRGCGTIGLGMGSDTGRLTYKFEHNLINRGYQAGAEIRLSEDKKVKLYA 632
Db 3588 TL-----EQARQLARKFQSTYEELTGM-----LREVEE-----LA 3628
Oy 633 TKPLSHPLNDQRLATLGYQOEVFGHSTNGFDLSTRTLEH-----EISRSIIONGGMN 684
Db 3629 TSGGSGPTGEOIPEFOQORKE-----LKEVMEHRLVLDIVNEVSRALLLVPR 3678
Oy 685 RTYSRLRLDKLKTQA 700
Db 3679 ---AREGLDKLVSDA 3690

RESULT 13
G6PD_STRPN STANDARD: PRT: 495 AA.
ID G6PD_STRPN 054537;
AC 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD).
GN ZMF OR SP1243.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RA Tetteijn H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radue D.,
RA Holtzapple E., Khouli H., Wolf A.M., Utechtack T.R., Hansen C.L.,
RA McDonald L.A., Feldblum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
RN [2]
RP SEQUENCE OF 281-433 FROM N.A.
RC STRAIN=ATCC 6323;
RA MEDLINE=98125733; Pubmed=9466257;
RA Coffey T.J., Enright M.C., Daniels M., Morona J.K., Morona R.,
RA Hyman-Litz W., Paton J.C., Spratt B.G.;
RT "Recombination exchanges at the capsular polysaccharide biosynthetic
RT locus lead to frequent serotype changes among natural isolates of
RT Streptococcus pneumoniae."
RL Mol. Microbiol. 27:73-83(1998).
RN [3]
RP SEQUENCE OF 281-433 FROM N.A.
RA MEDLINE=98192409; Pubmed=9533725;
RA Coffey T.J., Enright M.C., Daniels M., Wilkinson P., Berron S.,
RA Renoll A., Spratt B.G.;
RT "Serotype 19A variants of the Spanish serotype 23F multiresistant
RT clone of Streptococcus pneumoniae."
RL Microb. Drug Resist. 4:51-55(1998).
```

```
CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-
CC 1,5-lactone 6-phosphate + NADPH.
CC -1- PATHWAY: Pentose phosphate pathway; first step.
CC -1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
```

DR EMBL: AE007424; AAK75348.1; -

DR EMBL: 299802; CAB16927.1; -

DR EMBL: 299865; CAB16990.1; -

DR HSSP: P11411; IDPG.

DR TIGR: SP1243; -

DR InterPro: IPR001282; G6PD.

DR Pfam: PF00479; G6PD.1.

DR Pfam: PF02781; G6PD.C.1.

DR PRINTS: PR00079; G6PDHROGNASE.

DR ProDom: PD001129; G6PD.1.

DR TIGRFAMs: TIGR00871; zwf.1.

DR PROSITE: PS00069; G6P-DEHYDROGENASE; 1.

DR Oxidoreductase; NADP; Glucose metabolism; Complete proteome.

KW ACT_SITE 181

FT CONFLICT 348 348

FT V -> I (IN REF. 2 AND 3).

SEQUENCE 495 AA; 56790 MW; 99093BC9829FFAF CRC64;

Query Match 2.9%; Score 137.5; DB 1; Length 495;

Best Local Similarity 21.5%; Pred. No. 1.6;

Matches 125; Conservative 80; Mismatches 203; Indels 173; Gaps 33;

```
Oy 148 NSEGVVPTLEPEKGLYLR-LYARLFNCGVKNVPLAKKAFPOSSO-SEETASIGSHOK 205
Db 2 SSKVIV--TIIFASGDLMARKKILYPSLF-----RLTQSGMLSHFPAIGTARR- 46
Oy 206 TEPYANIKAALEDITQESAMDNGSIPIRLQALYAVARAAGYVDLSTIRNSIGEVVI 265
Db 47 --PWS--KEFFESVYVESLILDADSTEQAOEFA-----SHFYQO----- 81
Oy 266 IHDLGEPV-YIDYRAVEV---RGEGADKAFYVADEVPLLIGVFNHKKYETKNLIE 320
Db 82 SHDVNDSEHYIALRQLOALNEKYQAEHNKLF--FLSNAPQFFGTAKHLKSE---NIV- 135
Oy 321 NASAEHGFDDGWRMLRSVYVILPDNTADVSLIYDGTQYR-----FDEVEFTIDPKTN 374
Db 136 -----DGKGFRLI-VEKPFQTD-----YATASKLNDELATFDEQIFRIDHY-- 178
Oy 375 QLTTPDKLIPVKKRELLLEQLLVNMGAEAYNLQAVRALSNDLATRYFNWVNTIYVPERQ 434
Db 179 -----LGKEMIQSIFAVRA--NLIFENWNNDFID-----NQITAEIRIG 218
Oy 435 IQNDVSEFQSSSSTTEPAQVDESTLEPIETVELTDLGMDISPIEFSASNLIDOKLNL 494
Db 219 VEERGYYDQSGALRD---MVQNHLL-----QLLSLAMQ-KPASFYKDEIRAEKIV 267
Oy 495 VAAKARHLYDMDDVFLAINHDDGVNRSILGRISDAVSAYAAAILPDESENEVIDLPRT 554
Db 268 F-----KNLYH-PTDELEKHPKRGQYRS--GKI-DGMKTIYSTRSEPN----- 306
Oy 555 ALANRKPADVYOSKKVPLVYFVASDKPRDQIGLGMGSDTGRLVTFKFEH-NLINRDGY 613
Db 307 --VNESTTEFTTSGA-----FEVDSDRR--GVPPFFPRCKRLTEKGTNNIV----- 351
Oy 614 QAGAEIRLSEDKKVKVLYATRPPLNDQRLATLGYQOEVFGHSTNGFDLSTRTLEHEI 673
Db 352 -----FKQMSIFGEPLAPNI---LTIYIQPTGEFSLSLNKQYGE 389
Oy 674 SRSIIIONGGMNRTYSLRYRLDKIKTKQAPETWODLPVDFVN 714.
```

```

Db 390 EFNLAPN-----SLDYRTDATTATGASPEPEKLYIVLIN 423

RESULT 14
ID IF38-ARATH STANDARD: PRT: 900 AA.
AC 049160:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable eukaryotic translation initiation factor 3 subunit 8 (eIF3
pl10) (P105).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9065153; PubMed=9849901;
RA Kariol B., Yahalom A., Kwok S., Tsuge T., Matsui M., Deng X.-W.,
RA Chanowitz D.A.;
RT "The Arabidopsis homologue of an eIF3 complex subunit associates with
RT the COP complex." (1998).
RL FEBS Lett. 439:173-179(1998).
CC -1- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC METHIONYL-TRNAI AND MRNA (BY SIMILARITY).
CC -1- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EIF38 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF040102; AAC83464.1;
CC InterPro: IPR000717; PCI.
CC DR pfam: PF01399; PCI.1.
CC DR SMART: SM00088; PINT: 1.
CC Initiation factor; Protein biosynthesis.
SQ SEQUENCE 900 AA; 102989 MW; 3C03FEDBF59AB5D1 CRC64;

Query Match 2.9%; Score 137; DB 1; Length 900;
Best Local Similarity 20.1%; Pred. No. 4.3; Mismatches 252; Gaps 40;
Matches 156; Conservative 106; Indels 264;

OY 109 MSVIEETPLSLSELPFAQESTENGINPN-----DIPEYQGEOP 147
DB 270 LTKRIANTPKOKLEILFSVISAQFDVNRGLSGHPIPWKKCVLMLTILDIIVKY----- 324
OY 148 NSEVVVPPLLEPE-----KP-----GLIKRLYARLFNDGVKVRPLAKFYQSSQSGTSA 198
DB 325 -SNIVVDVDEPEDENETSKPTDYDKI-----RYMGLVAFLEBRVDEFEFKSLQC----- 373
OY 199 IGSSHKTEPRYAKALADITDES-----AMLNGSIPPL---RQTALVAARVG--YY 248
DB 374 -----IDPHT--REYVELLRDEPMFLALQNIODYFEMRGDFKAAKVALRREVAIYY 424
OY 249 DIDLSIRNSIGEVNIIHDLGEPYIDYRAVEVRGSGADKAFITVADVP----- 300
DB 425 KQO-----EYDAMKRLALVEEBEETLEAKESGPTSTFYVPEVVPKRPFPES 475
OY 301 -----LLIGDVFHNKGYETKKNLINASAEHGYFDGRWLDRSVDVILPDNTADVSLIYD 354
DB 476 SRAMMDILVLIYRNGDERKARAMLCDINHMLNDNFYAR--DLLMSHLDNNTQHMD 533
OY 355 TGTQYVFEDEV-----FTTIDPKTNQLTTDDKLPYKRELLEQLLTVMGSAVNIQAV 407
DB 534 ISTQILFNRTMOLGICAF-----RAGMITESHSL-----SELYFGORV 573

```

```

OY 408 R-ALSNDLIATRYFNMYNTEIPEPEROIONDOVSFEOSSSSRTEPAQVDESTLEPIET 466
DB 574 RELLAQGVQSQRHEKT-----PEQERMER-----RQMPYHM-ILNE-LLEA 615
OY 467 VELTDGLMDISPIERSASNLIDQKLVAAKARHLYDMEDDRVLAINHDDGVRSILGR 526
DB 616 VHLICAMILEVPMN---AANSHDKRRVISKNFRLLEISROAFTAPEN----- 663
OY 527 ISDAVSVAARILPDESEN--EVID-LPERTALANRKTADV-----YQSKVPLVYVAS 579
DB 664 VRDHVMAATRALTKGDKQKAFVNLNSLEWRLLKNRSIILDMVKDRQVEEAVRTYLETYS 723
OY 580 DKPRDQIGLWGSQDGTGRTLVTFEHLINRDGYQAQAEILRSEDKGVLYATKPLSPH 639
DB 724 -----SSYES-----LSLDQAKMFVDSEPVHS 747
OY 640 -----LNDQLRATLGYQOEVEFGHSTNGFDLSTFLE-HEISRSIIONGNNRTYSLRYR 692
DB 748 IVSKMAMINEELHAS-----WDQPTFCIVFEVQHSRLQ-----SLAFQ 785
OY 693 L-DKLTQAPPEPTWQDLPVDFVNGKPSQEALLAGVAHKTVAOVLVPMGRQRYG-LE 750
DB 786 LTERKSLIAE-----SNERAMESRTGGGI-----DLSRRRDNODVYGA 827
OY 751 VGSSGLVSD-ANMAIA---RAGISGVYSGDNAY--GSNRRAHQMTGIGQGYIWSDN 801
DB 828 SGGGGMQDANQAGCGQGRSGVGGGRSSGQNGMGSQNR-----GGGYAGRGVSGSN 880

RESULT 15
TYCC_BACBR
ID TYCC_BACBR STANDARD: PRT: 6486 AA.
AC 030409:
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrocidine synthetase III (includes: ATP-dependent
DE adenylyase (asn) (Asparagine activase); ATP-dependent glutamine
DE adenylyase (gln) (Glutamine activase); ATP-dependent tyrosine
DE adenylyase (tyr) (Tyrosine activase); ATP-dependent valine adenylyase
DE (Val) (Valine activase); ATP-dependent ornithine adenylyase (orn)
DE (Ornithine activase); ATP-dependent leucine adenylyase (Leu) (Leucine
DE activase)].
GN TYCC.
OS Bacillus brevis.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX NCBI_TaxID=1393;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 8185;
RX MEDLINE=98012987; PubMed=9352938;
RX Klotz H.D., Marahiel M.A.;
RT "The tyrocidine biosynthesis operon of Bacillus brevis: Complete
RT nucleotide sequence and biochemical characterization of functional
RT internal adenylation domains."
RL J. Bacteriol. 179:6843-6850(1997).
CC -1- FUNCTION: INCORPORATES SIX AMINO ACIDS (FOR TYROCIDINE A, ASN,
CC GLN, TYR, VAL, ORN, AND LEU) IN THEIR L-CONFIGURATION INTO THE
CC PEPTIDE PRODUCT.
CC -1- COFACTOR: CONTAINS 6 COVALENTLY BOUND PHOSPHOPANTETHEINES (BY
CC SIMILARITY).
CC -1- PATHWAY: Cyclic peptide antibiotic tyrocidine biosynthesis.
CC -1- SUBUNIT: LARGE MULTENZYMIC COMPLEX OF TYCA, TYCB AND TYCC.
CC -1- DOMAIN: CONSISTS OF SIX MODULES, AND HARBORS A PUTATIVE
CC THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE
CC INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
CC FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
CC ADENYLATION, THIOYLATION, CONDENSATION (NOT FOR THE INITIATION
CC MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
CC (OPTIONAL).
CC -1- MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC
CC DECAPEPTIDES, TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-

```

```
CC ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR
CC RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
CC CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 6 ACYL CARRIER DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF004835; AAC45930.1; -.
DR HSSP: P14687; 1AMU.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR001242; Condensatn.
DR InterPro: IPR003880; Ppantne_attach.
DR InterPro: IPR00379; Ser_estrs_site.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00501; AMP-binding; 6.
DR Pfam: PF00550; pp-binding; 6.
DR Pfam: PF00668; Condensation; 6.
DR Pfam: PF00975; Thioesterase; 1.
DR PRINTS: PR00134; AMPBINDING.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 6.
DR PROSITE: PS00455; AMP_BINDING; 6.
DR PROSITE: PS50075; ACP_DOMAIN; 6.
KW Ligase: Antibiotic biosynthesis: Phosphopantetheine;
KM Multifunctional enzyme: Repeat.
FT REPEAT 466 1038 DOMAIN 1 (ASPARAGINE-ACTIVATING).
FT REPEAT 1521 2070 DOMAIN 2 (GLUTAMINE-ACTIVATING).
FT REPEAT 2536 3113 DOMAIN 3 (TYROSINE-ACTIVATING).
FT REPEAT 3590 4149 DOMAIN 4 (VALINE-ACTIVATING).
FT REPEAT 4606 5203 DOMAIN 5 (ORNITHINE-ACTIVATING).
FT REPEAT 5658 6245 DOMAIN 6 (LEUCINE-ACTIVATING).
FT DOMAIN 970 1037 ACYL CARRIER (ACP) 1.
FT DOMAIN 2007 2074 ACYL CARRIER (ACP) 2.
FT DOMAIN 3045 3112 ACYL CARRIER (ACP) 3.
FT DOMAIN 4080 4147 ACYL CARRIER (ACP) 4.
FT DOMAIN 5124 5191 ACYL CARRIER (ACP) 5.
FT DOMAIN 6167 6234 ACYL CARRIER (ACP) 6.
FT BINDING 1000 1000 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 2037 2037 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 3075 3075 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 4110 4110 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 5154 5154 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 6197 6197 PHOSPHOPANTHETHEINE (BY SIMILARITY).
SQ SEQUENCE 6486 AA: 724011 MW: 4934900AF07DF786 CRC64:

Query Match 2.98; Score 136; DB 1; Length 6486;
Best Local Similarity 18.5%; Pred. No. 1e+02;
Matches 139; Conservative 102; Mismatches 285; Indels 226; Gaps 33;

OY 44 AHDAIINQAKGNPNVLTPEQIOARLNAAGL--NAKP-----OSOA 83
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 269 ADDVVFSSVSVSGRPP-----QIOGIESMVGILFTIPTRVQTNKOOTFSELLOTVQKOA 322
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 84 LDVNVFP-----DOSPIRIGOSGPPGLG----- 107
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 323 LASATYFAPFLYEIQSTFTVLKQELIDHLVTFENYPDHSMKMLESLGQTFVESGDEQTS 382
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 108 -DMSVIEETTPLSLELFAOSTEMGINPDNYIEPYGEOGRN--SEVVVPTLEPEKFG 164
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 383 YDLNVVVALPASN--ELYVKLSYNAAYESSFVNRIECHLFTVIDQYIGNPHVHLHEIGI 440
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 165 I-----KRLYARLFNDGNKVPRLKAKFYQSSQSGEISAIGSHOKTEPYANIKALEIT 220
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 441 ITEEEKQQLLVAVYDITAEYPRDKTIF-----ELIAQASRTPAKKAIVGCCEDTLT 491
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 221 QESAMDLNGSIPR-LRQTALVAARAVGY--YDIDL-----SIIRNSIGEVVITIHDLGE 271
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 492 YQELMERASQULANALREKGIASGTSIVSAEHSLELYAIMAVLSSGAAYLP-----D 545
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 272 PYIDYRAVEVYEGADAKAFETTVADEVPL-----LIGVFHNGKYEYRKNLINENASAEH 326
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 546 PEYPODRIRQYL-----LDDSQTLILLQSHLPQNPINFAGSVL-----YLDDRSLEYEGGSTS 596
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 327 GYEDGRWLDRSVYVILLPONTAD--VSLIYDGTGYRFDVEVFEFTIDPKTNQLTTPDPKLP 384
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 597 -----FAPESKPPDLLAYMYTSSGS-----TGNPKGAMIT----- 625
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 365 VKRELLEDQLLTYN-----MGEAYNLQAVRALNSDLATRFENNV-----NNEIYPEREEOIN 437
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 626 -HQGLVNYIMWANKVYVGEAEVDPPLYSISFDLVTSIFPFLSGNTIHVRGADKVO- 683
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 438 DQVSFEQSSSRTERPAQVDESTLEFVETVELTDLGILMDISPIEFSASNL--IDOKLNL 494
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 664 -----VILDIIDKQNKVGIIKLTPLHKLIEHIDGKASSIRRFIVGGENL 727
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 495 VAAKARHLXDMPPDRVLAINHDDGVNRSILGRISDAVSAVARAILPDESENEYIDLPERT 554
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 728 PTKLAKQIYDHGEGVQJFNE-----YGPTEYVVGCMIVLYDPOTTQDSVPI----- 775
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 555 ALANKRTPADVYQSKKVPPLYVFAVDKPRD-GQIGLGMSDPTGFRLYVFEHNLIINRDX 613
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 776 -----GVPRD-----NVQLYLDASMKQVPVPSGLG-----EMTYAGDGV 809
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 614 QAGAELR--LSEDK-----KGVKLYATKPLSHPLND--QLRLTGLYQGVDFGSHSTNG 661
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 810 AKGYNRBELTKEKFTIDNPFPGTKMYRTGDGLAKMLPDGMEMYAGMDYQVKIRGHRIEM 869
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 662 FDLSTRITLEHETSRII-----QNGGMNRTYS 688
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 870 GEITRILTQHEAVKKAIVYIVKDESGQNVLYA 901
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 16
ID BCSC_PSEFL STANDARD: PRT: 1279 AA.
AC P58937;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose synthase operon protein C precursor.
GN BCSC OR W5SE.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SRW25;
RA MEDLINE=22013850; PubMed=12019221;
RX Spiers A.J., Kahn S.G., Bohannon J., Traviano M., Rainey P.B.;
RT "Adaptive divergence in experimental populations of Pseudomonas
RT fluorescens. I. Genetic and phenotypic bases of wriukly spreader
RT fitness".
RL Genetics 161:33-46(2002).
CC -1- FUNCTION: Required for maximal bacterial cellulose synthesis (By
CC similarity).
CC -1- PATHWAY: Bacterial cellulose biosynthesis.
CC -1- SIMILARITY: CONTAINS 10 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AY074776; AAL71845.1; -.
KW Cellulose biosynthesis; Repeat; tpr repeat; signal.
```

```

FT SIGNAL 1 21 POTENTIAL
CHAIN 22 1279 CELLULOSE SYNTHASE OPERON PROTEIN C.
REPEAT 27 60 TPR 1.
REPEAT 62 94 TPR 2.
REPEAT 218 250 TPR 3.
REPEAT 306 339 TPR 4.
REPEAT 384 417 TPR 5.
REPEAT 460 493 TPR 6.
REPEAT 495 527 TPR 7.
REPEAT 606 639 TPR 8.
REPEAT 719 752 TPR 9.
REPEAT 787 820 TPR 10.
SEQUENCE 1279 AA: 138664 MM: EAYE42BE7810D201 CRC64:

Query Match 2.8%; Score 134.5; DB 1; Length 1279;
Best Local Similarity 19.1%; Pred. No. 10;
Matches 199; Conservative 140; Mismatches 334; Indels 369; Gaps 54;

QY 28 ALAQONNP-----ANITNHVPAHDTAINQAKGNPVLVLT-----TPEDIQARLNAGLNAK 78
DB 374 AQAERLNPGPGGAIALAGFOADQNPDDAEAGYRKYLARHPGDPDALSLGLINVLSSQSG 433
QY 79 POSOALDVNFDDQSPISRISEOSPLGLMSVIEETTPLSLEELFQOESTEMGINNDY 138
DB 434 P-DEALKLI--DVSFPAQR-AKRAPSVKIN-----ALRAVOVG-----K 468
QY 139 IPEYQEQPSEVVPPTL--EPEKPG---LIRRLYARLFNDGVNKPRLKAFYQSSOS 193
DB 469 LARGQGLKAAQAAYRQALADPENPWRALARMYLR---DQIRARALIDGLKQSP 525
QY 194 GFTSAIGSSHQKTEPYANIKAALEDITQESAMDLNGSIPRLQRTALVAARAYYDIDS 253
DB 526 NQDPAIYTS-----TILSAQJSEMKAQEA--TLGRIPTAQRTA-----DMN 564
QY 254 IIRNSIGENVVITHDGEPIYIDYRAVEYEGADKAFITVADDEVLLIGDVFHKKYE 313
DB 565 EL-----AIDIALHQOTD-----IALETARRGPRPALALLGRSEPLT-----RNKPE 607
QY 314 TKKNL-----TENASAEHGYPGDRMLDRSDVI--LPDNTADVSLIY----- 353
DB 608 RVAVLAAAYVEGAQYG-----LDMNQKVENNPNPTDQKLLYANVLLKANKYSEAG 661
QY 354 -----DTGTOYRDEVEVFFITDPKTNQLTTPDPKLPVKRELLBOLLTVNKG 401
DB 662 ELIREVOGQPLTFETGRQ-RYDDLILY----- 686
QY 402 YNLQAVRAL--SNDLITATREYNVNTETVPEREQIONDOVS-----PEQSSSSTEP 452
DB 687 YRKQADALREKNDLVAA--YDMLIS-----PALAQRPNDAIGVALARMYAAASNGK-- 736
QY 453 AQVDESTLEPVIETVELTDGILMDISPIEFSASNLIDQKLNL--VAAKA--RHLYDKPDD 508
DB 737 -----KAMEL-----YAPLIOQNNRNARLOGLADIALKGDRGLAQSASD 777
QY 509 RVLAIHHDGVNRSI-----LGRISDAV-----SAVARAILPDE 542
DB 778 KALAL-----EFGNBEILTSARIIYQGLKGNSEAEALLKALAIENAMKAKYOVAQASAPGT 834
QY 543 SENEVIDLPERTALANRKTTPADYQSKKVPVLYFVASDKRQDQIGIGKMSDGTRLVTK 602
DB 835 STNPFVGLP-----GORROVTDLTGAVAVPPPT--DAPTKSVTSNAFASATSNLSDP 885
QY 603 F-----EHNLIINRDGYQA-GAELTSEDKKGVK---LYATKPL 636
DB 886 FVPPSSIASIDSPELSPARALDTILDRGRGYVVGGLSVASNGGKLSITTVAEAFEA 945
QY 637 SHPLNDQ---LRAT----- 650
DB 946 RHPVGVNTVALRPTPVHLSAGSVKAESSLRFKGTGTEPAGSQSDSGVGLAVAFENPDQGL 1005
QY 651 OOEVEFCHSTNGPDLSTRTLEHLSRSIIONGS--WNRTYSLARTLDKLIKQAPPE----- 703
DB 1006 KADV--GVSPGLGLFYNTLVGCVSVSRPEFANSNRPYGANISRRPVTDSVNSFAGSEDCAGN 1064

```

```

QY 704 TWODLPVDFVNGKPSOEALLAGV-----AVHKTVAIDLNVNPMRGYRQRYSLFVGS--GLV 757
DB 1065 KMGVTTANNGRGELSYDNQKAGYGVASLHELLGNVNDNR-----LELGSGIYWL 1117
QY 758 SDANMAIARAGISG-----VSFQDNAYGSNRAHQMTGTGIQAGYIWSDFNHPVY 807
DB 1118 RNNPRDTLTLIGISGSAMTFKEMODEFYTGNGCYFSPQFFSLG---VPIRMAQSFDRFSY 1174
QY 808 RLRFAG-----GDOSIRGADHSLSPISKGYLTGGQVLAFTAEYNEFM 854
DB 1175 QVKS SVGLQHTAGCADYFPEDSTLQATKN--NPKYDSTSKTG-----VG---YSFNAA 1223
QY 855 KDLRLA-----VFGDIG--NAYD 870
DB 1224 AEYLLSSRFYLGELIGLDMQD 1245

RESULT 17
P200_MYCPN STANDARD; PRT; 1036 AA.
ID P200_MYCPN
AC P75211: 050346;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein P200.
GN P200 OR MPN567 OR MP275.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=96257197; PubMed=8675035;
RA Proft T., Hilbert H., Plagens H., Hermann R.;
RT "The P200 protein of Mycoplasma pneumoniae shows common features with the cytochrome-c-associated proteins Hmw1 and Hmw3."
RL Gene 171:79-82(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreisch R., Hilbert H., Plagens H., Pirkl E., Li B.-C., Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
RN [3]
RP SEQUENCE OF 641-678 FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=95075318; PubMed=7984111;
RA Proft T., Hermann R.;
RT "Identification and characterization of hitherto unknown Mycoplasma pneumoniae proteins."
RL Mol. Microbiol. 13:337-348(1994).
CC -1- FUNCTION: PROTEIN CYTOSKELETON-ASSOCIATED WHICH COULD BE AN ACCESSORY STRUCTURAL COMPONENT IN CYTADHERENCE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL: U25989; AAC99815.1; -
CC DR EMBL: AE000027; AAB95923.1; -
CC DR EMBL: Z32646; CA83569.1; -
CC KW Cytoadherence; Structural protein; Repeat; Complete proteome.
CC FT DOMAIN 277 280 POLY-PHR.
CC FT DOMAIN 300 845 PRO-RICH.
CC FT DOMAIN 357 360 POLY-PHR.

```


DB 622 KQIGRPTFD-AEKVTIDLSKVYVQADASLNEKDIAVAANAISGRKADVTASTDLHLQ 680
QY 510 -----VLAINHDDGVNRSILGRISDAVSAYARALP-----DESENEVIDL----- 550
DB 681 NTVTAMKTNADSV---VL-----AMGYKYLLVLPFVKNVGEDENTAVQLTNGSETV 732
QY 551 -----PERALANM-TPADY-YQSKVPL-----YFVASDKPRD-GQIGLG 590
DB 733 TWTVINHVGSNSKDVKNCKTGVSVSLHMDIPLOTIKIYEVSSERPRANYGGITEE 792
QY 591 WGSDD-----TGFRLYTKFEHNLIIRD-----GYOAGAEIR---LSBDKKGVLYATKPL 636
DB 793 KGMNDVLDLTHTRFTGKW-HAIIINYLKVGDKTLKAGTDSAYILLEND-----NKDL 845
QY 637 SHPLNDOLRATLGYOQEVFGHSTNGFDLSTRLEHISRSIIIONGWNRTYSRLYRDLK 696
DB 846 TETMNQALLAAL-----NEGSNKYKQAMSYYLEVER---ITGQVENQTNTYKELY 896
QY 697 KTOAPEPTMODLPVDFVNGKPSQOEALLAGVAVHKYVADNLVNMKRGYRORYSLEVGSSGL 756
DB 897 RNTVVTHTPDP-----KPTK-----AVHNKKGEDL----- 923
QY 757 VSDANAIARAGISGVYSPGDNAVGSNRAHOMTGIOAGYIMSDNHNHYRLRFFAGD 816
DB 924 -----NHGKYARGDVLSTYEMTWDLK-GYDKFAEDTYD---LATGV 960
QY 817 QSRGYANDLSLSPISD-----KGYLT-----GGQVLAVGTA 847
DB 961 SFDDYDETKVPPIKDLAVKDSKSGEDINTQFTISWDKAGTYTISAKDPOAFILAHGQ 1020
QY 848 EYNTEPMKDLRLAVGDCIDN-AYDKGFTNDTKIGAGVG-VRRNASPGQVRVDVATGVKEE 905
DB 1021 ELRVTLPTVKKANQSDVYNLALEONFQGRIKTNTVYVNHPIKVPKQDVYIKVGRQSON 1080
QY 906 GNPRL 911
DB 1081 GATIKL 1086

RESULT 19
FAT2_HUMAN
ID FAT2_HUMAN STANDARD: PRT: 4349 AA.
AC Q9NYQ8: Q75091: Q9NSR7:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protocadherin Fat 2 precursor (hfat2) (Multiple epidermal growth factor-like domains 1).
GN FAT2 OR MEGF1 OR CDHF8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20202599; PubMed=10716726;
RA Wu Q., Maniatis T.;
RT "Large exons encoding multiple ectodomains are a characteristic feature of protocadherin genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
RN [2]
RP SEQUENCE OF 3777-4349 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
RN [3]
RP SEQUENCE OF 4142-4349 FROM N.A.
RC TISSUE=Testis;
RA Pouster A., Wellenreuther R., Mewes H.-W., Well B., Wilmann S.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 32 CADHERIN DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF231022; AAF61928.1; -
CC EMBL: AB011535; BAA32463.1; -
CC EMBL: AL157443; CAB75663.1; -
CC Genew: HGNC:3596; FAT2.
CC MIM: 604269; -
CC HSSP: P00740; IEDM.
CC InterPro: IPR002126; Cadherin.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR001791; Laminin_G.
CC Pfam: PF00008; cadherin; 33.
CC Pfam: PF00008; EGF; 2.
CC SMART: SM00205; CADHERIN.
CC SMART: SM00181; EGF; 2.
CC SMART: SM00282; LamG; 1.
CC PROSITE: PS00232; CADHERIN_1; 14.
CC PROSITE: PS00268; CADHERIN_2; 32.
CC PROSITE: PS00222; EGF_1; 2.
CC PROSITE: PS01186; EGF_2; 1.
CC PROSITE: PS00025; LamG_DOMAIN; 1.
CC KW Transmembrane; Glycoprotein; Repeat; EGF-like domain; Signal.
CC FT SIGNAL 1 18
CC FT CHAIN 19 4349
CC FT DOMAIN 19 4048
CC FT TRANSMEM 4049 4069
CC FT DOMAIN 4070 4349
CC FT DOMAIN 34 148
CC FT DOMAIN 149 256
CC FT DOMAIN 363 458
CC FT DOMAIN 459 564
CC FT DOMAIN 565 669
CC FT DOMAIN 716 820
CC FT DOMAIN 821 925
CC FT DOMAIN 926 1032
CC FT DOMAIN 1033 1137
CC FT DOMAIN 1138 1242
CC FT DOMAIN 1243 1346
CC FT DOMAIN 1350 1448
CC FT DOMAIN 1449 1555
CC FT DOMAIN 1556 1660
CC FT DOMAIN 1661 1758
CC FT DOMAIN 1759 1872
CC FT DOMAIN 1872 2070
CC FT DOMAIN 2071 2171
CC FT DOMAIN 2172 2272
CC FT DOMAIN 2273 2379
CC FT DOMAIN 2380 2481
CC FT DOMAIN 2482 2585
CC FT DOMAIN 2586 2691
CC FT DOMAIN 2692 2797
CC FT DOMAIN 2798 2906
CC FT DOMAIN 2907 3011
CC FT DOMAIN 3012 3113
CC FT DOMAIN 3114 3218
CC FT DOMAIN 3219 3321
CC FT DOMAIN 3322 3426
CC FT DOMAIN 3427 3531
CC FT DOMAIN 3532 3642
CC FT DOMAIN 3773 3944
CC FT DOMAIN 3951 3984
CC FT DOMAIN 3989 4022
CC -----
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).
CC CADHERIN 1.
CC CADHERIN 2.
CC CADHERIN 3.
CC CADHERIN 4.
CC CADHERIN 5.
CC CADHERIN 6.
CC CADHERIN 7.
CC CADHERIN 8.
CC CADHERIN 9.
CC CADHERIN 10.
CC CADHERIN 11.
CC CADHERIN 12.
CC CADHERIN 13.
CC CADHERIN 14.
CC CADHERIN 15.
CC CADHERIN 16.
CC CADHERIN 17.
CC CADHERIN 18.
CC CADHERIN 19.
CC CADHERIN 20.
CC CADHERIN 21.
CC CADHERIN 22.
CC CADHERIN 23.
CC CADHERIN 24.
CC CADHERIN 25.
CC CADHERIN 26.
CC CADHERIN 27.
CC CADHERIN 28.
CC CADHERIN 29.
CC CADHERIN 30.
CC CADHERIN 31.
CC CADHERIN 32.
CC LAMININ G-LIKE.
CC EGF-LIKE 1.
CC EGF-LIKE 2.

```
FT DOMAIN 3297 3302 POLY-SER.
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 459 459 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 789 789 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 966 966 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1175 1175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1303 1303 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1383 1383 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1417 1417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1904 1904 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1998 1998 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2007 2007 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2165 2165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2183 2183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2325 2325 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2368 2368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2387 2387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2430 2430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2470 2470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2547 2547 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2597 2597 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2654 2654 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3125 3125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3276 3276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3310 3310 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3430 3430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3471 3471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3601 3601 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3772 3772 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3813 3813 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3840 3840 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3873 3873 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3904 3904 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3989 3989 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4117 4117 P -> L (IN REF. 2).
FT CONFLICT 4160 E -> G (IN REF. 2).
FT SEQUENCE 4349 AA: 479383 MW: 6C050CA17AD398ED CRC64;
```

Query Match 2.88; Score 133.5; DB 1; Length 4349;
Best Local Similarity 17.48; Pred. No. 75;
Matches 199; Conservative 162; Mismatches 388; Indels 395; Gaps 50;

```
OY 52 AKAGNPVLLTPED--IQARLNLAKLNAKPOSOALDVNFDOSPIRIGESPP-LGLD 108
DB 2145 ARDGGPFSLOSSEEVLLTVTNKSNPLFOSPYKYRVPENITLTYPIHLHTQKRSFGLRLI 2204
OY 109 MSYIEE-----TTPLSLE---ELFAQESTENGINPNDIPEYGEOP 147
DB 2205 YNIVEEERPLMLFTTDFKTVGLTVGPLDYESKTKHVFVTRATDNLG-----SFESEA 2256
OY 148 NSEVVV-----PPTLE-----PEKRLIKRLVA-----RLFNDGV 177
DB 2257 TVEVLEEDVDNPNPTSQLVYTTTSSISGLPAQTPFVLIQLLSDDOSGRNRDVSQYIVEDGS 2316
OY 178 NKVPRLLAKAFYQ--SOSGETSAIGSHQKTEPYANIKAALED-----ITQESAM-----D 226
DB 2317 D-----VSKRFQJNGSTGEMSTYQELDYEAQOHFHVKRAMDKDPLLTGTLTLLVVAVSD 2371
OY 227 LNSIPRLKOTALVARAGYVYIDLSIINNSIGEVDVITHDGEPIYIDYRAVEVGE- 285
DB 2372 INDNPEFRQPOYEANVS-----ELATCGHLVLYKQALIDPSDSRTSRLELYLLSGNDR 2425
OY 286 -----GADKAFITVADEVYLLIGDV-----FH 308
DB 2426 HFTINSSGSIISFNPLCKKLLDSYNLRVAGASGVFRAT--VFVYNTTNANKYSPEFQ 2482
OY 309 HGKYEKKKLIENASAEHGYFDGRWLDKRSVDVILLPONTADVSLIYDTGTQYRDEVEVFT 368
```

```
DB 2483 QHLYAE--LAENAVGTRKVIDLLAIDKDSG--PYGTIDYTIINKLASEK-----FS 2530
OY 369 IDPKINOLTTTPDKLPVYRRELLEOLLYVMGEAYMLQAVRLASND----- 413
DB 2531 INP-NGQIAT-----LQKLDRENSTE--RVIAIKVARGCGRAVACTYKIILT 2576
OY 414 -----LIATRYENMYNT-----EIVPEREQIONQVSEFQSSSRTEPAQVD 456
DB 2577 DENNDPQPFKASEYVLSQSNVSKDSPYIQLVAYADAGEQNDADVY-----SVNPEDLYKD 2632
OY 457 ESTLEPYIETVELTDCIL-MDISPIEFSA-----SNLIQKMLVAAKAR----- 500
DB 2633 VLEINPVYGVVKKVKSJLQGLEQTLDFEIKADQDGPPIHNSLVPRQLQVPPKVSLEPKRS 2692
OY 501 ---HLYPDPPDRLVLAINDDGVNRSILGRISDAVSAVARAILPDSESENEVIDLPERTALA 557
DB 2693 EPLYTFSAPED--LPEGSEIGIVKAVAAQ--DPVYISLVKGTTPESNKKQGVESLDDPTGYI 2749
OY 558 NRKTPAD-----VYQ-----SKK 570
DB 2750 KYRKPMDESTRKLYQIDVMAHCLQNTDVVSLVSNIQVGDVNDNRPVEADPYKAVLTEN 2809
OY 571 VPL-----YVFVASDKRRDQIGLGMSDTCRLVTFEHLINRD----- 611
DB 2810 MPVGTSVIQTVAIDKTDGRDQGVSYRLSDPGSNV-----HELFAIDSESGWITTLQELD 2864
OY 612 -----GYQAGAEELRLS-----EDKKGVKL 630
DB 2865 CETQCTYHFHVAYADHGTQLQSALQVSTTDENDNAPRASREYRGSVENSEPEL 2924
OY 631 YAT-KPLSHPLNDQLRATLGIQOE-----VEGHSINGFD--LSTRTLBEHISRSIIONG 681
DB 2925 VATLKTLDADISEQNRQVTCYITTEGDPILGQFGISQVGEWIRISSRKTLDREHTAK----- 2979
OY 682 GMRNRYSLRYLHDKKTKQAPPEPMODELPVDFYNGKPSQSEALLAGVAKRYADNLVNMFR 741
DB 2980 -----YLLRVTASDCKFOA-SYVEIEFVLDVNDNSPQCSQLLYTGKVEDV----- 3024
OY 742 GYRORYSLEVSSGSLVSDANNAIARAGISGVSEFGDNNVGSNRHQMGIGIOAGYWS-- 799
DB 3025 -FPGHFLIKVASATLDDTDITNOI-----TYSL-----HGPG-AHEFKLDPHTGELTLLT 3071
OY 800 -----DNFNHVPRLRLEFPFAGGDSITRGVYAHDSLSPISDKCYLTGGOVLAVGTAEVYE 852
DB 3072 ALDRERKQVFNLY--AKATDGGGRCQADITLHVEDVNDNA-----PRF 3113
OY 853 FMKDLRLAVFGD--IGNAYDKGFTNDRKIGAGVGRMASP--VGQVADVATGVKEG 907
DB 3114 FPRSHCAVAFNNTYKTVAVAVVAFARDPDQGANNOVYVSLPDSABGHESIDATGTYIRLEK 3173
OY 908 PIKL 911
DB 3174 PLQV 3177
RESULT 20
TRX_DROME STANDARD; PRT; 3726 AA.
AC P20659; Q27255; Q27327;
DT 01-FEB-1991 (Rel. 17, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ttthorax protein.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90192757; PubMed=2107543;
```

RA Mazo A.M., Huang D.-H., Mozer B.A., David I.B.;
RT "The trithorax gene, a trans-acting regulator of the bithorax complex
in Drosophila, encodes a protein with zinc-binding domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2112-2116(1990).
RN [2]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RA MEDLINE=95009521; PubMed=7924996;
RX Sedkov Y., Tiliib S., Mizrokh L., Mazo A.;
RT "The bithorax complex is regulated by trithorax earlier during
Drosophila embryogenesis than is the Antennapedia complex, correlating
with a bithorax-like expression pattern of distinct early trithorax
transcripts.";
RL Development 120:1907-1917(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=Oregon-R;
RL MEDLINE=96100387; PubMed=8555104;
RA Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95047388; PubMed=7958911;
RA Kuzin B., Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "The Drosophila trithorax gene encodes a chromosomal protein and
directly regulates the region-specific homeotic gene fork head.";
RL Genes Dev. 8:2478-2490(1994).
CC -1- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- MISCELLANEOUS: THIS PROTEIN HAS BEEN EXPERIMENTALLY SHOWN TO BIND
ZINC.
CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -1- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; M31617; AAA29025.1; -;
DR EMBL; Z50152; CAA90514.1; -;
DR EMBL; Z50152; CAA90513.1; -;
DR EMBL; Z31725; CAA83516.1; -;
DR EMBL; Z31725; CAA83515.1; -;
DR PIR; A35085; A35085.
DR HSSP; P20393; 1A6T.
DR TRANSFAC; T00850; -;
DR Flybase; FBgn0003862; trx.
DR InterPro; IPR003889; FYRICH_C.
DR InterPro; IPR003888; FYRICH_N.
DR InterPro; IPR003616; PostSET.
DR InterPro; IPR001214; SET.
DR InterPro; IPR001965; Znf_PHD.
DR InterPro; IPR001841; Znf_Ling.
DR Pfam; PF00628; PHD; 3.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00542; FYRC; 1.
DR SMART; SM00541; FYRN; 1.
DR SMART; SM00249; PHD; 4.
DR SMART; SM00508; PostSET; 1.
DR SMART; SM00184; RING; 3.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS50280; SET; 1.
DR PROSITE; PS01359; ZF_PHD_1; 4.

DR PROSITE; PS50016; ZF_PHD_2; 3.
KW Transcription regulation; zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Developmental protein; Activator;
KW Alternative splicing.
FT ZN_FING 1266 1347 PHD-TYPE 1.
FT ZN_FING 1348 1393 PHD-TYPE 2.
FT ZN_FING 1421 1482 PHD-TYPE 3. (ATYPICAL).
FT ZN_FING 1734 1793 PHD-TYPE 4 (ATYPICAL).
FT ZN_FING 1794 1844 PHD-TYPE 5 (ATYPICAL).
FT DOMAIN 3599 3708 SET.
FT DOMAIN 512 516 POLY-SER.
FT DOMAIN 565 570 POLY-ASP.
FT DOMAIN 661 664 POLY-SER.
FT DOMAIN 905 910 POLY-SER.
FT DOMAIN 1576 1582 POLY-SER.
FT DOMAIN 2298 3027 GLN-RICH.
FT DOMAIN 3032 3040 POLY-SER.
FT DOMAIN 3181 3184 POLY-GLN.
FT DOMAIN 3220 3225 POLY-GLU.
FT VARSPLIC 1 368 MISSING (IN SHORT ISOFORM).
FT CONFLICT 2025 P -> PWLTSPLEFLGISTHGILLWLLGGVVRKGG
(IN REF. 1).
FT CONFLICT 2341 2341 R -> S (IN REF. 1).
FT CONFLICT 2392 2392 G -> S (IN REF. 1).
SQ SEQUENCE 3726 AA; 400575 MW; D2756E50763D1CF5 CRC64;

Query Match 2.8%; Score 133; DB 1; Length 3726;
Best local Similarity 18.3%; Pred. No. 63;
Matches 156; Conservative 122; Mismatches 281; Indels 292; Gaps 43;

QY 13 MPVALAAYPLMTS-----QALAQNNPANIINHVAHDATINQAKAGNPVLLTPEQI 66
DB 2397 MPITTLAEAVVQSOFVMEPOALEQ--ELANRVQHFSTSSSS--SSNCSLPTNVVNMQ 3054
QY 67 QARLNAGLNKAPQSQLDVNVNPDQSPISR---IEQSPRLDMSVIEETPLLEE 122
DB 3055 QABSTTSSSTRPTNRKPLMQQROEPAPLSNECPVSSPTPKPEQPIIHQMTSASVSK 3114
QY 123 LFAQESTEMKINPDYIPREYQGEQPNSEV---VVP---PTLEPEKGLIKRLARLFND 175
DB 3115 CYAKKST---LPS---PYVEAEIKVSSVLESITPDTMAILEQ---VQSITYE 3162
QY 176 GVNVKVPRLAKKTYQSSQSGETSAIGSHQTEPYANIKALEDTQGSAMDLSIDPLRL 235
DB 3163 GL-----YEKNPGRGS-----KTEQLLLQQQREQLNQ--LVNNG----- 3196
QY 236 QTAIVARAAGYDIDLSIIRNSTIGEVDYIHDGEVYIDRAVEVRGEGADQKAF--- 292
DB 3197 -----YLLDKHTTQVERPMDPV-----YREEDLEEEDDDDFSLK 3232
QY 293 --TTVADEVPLLAGDVHNGKYEKKMLIENASAEHGYPDGRWLDRSVDVLLPNTADV 350
DB 3233 MKTSACND-----HEMSDEEPVAVDKDKS--ILDNLTDDCADSATATITHEVD 3280
QY 351 LIYDTGTQVRFDEVEVFTIDPKTNLTTPDKLPVKRELLQLLTVNNGEAYNLQAVRAL 410
DB 3281 A--SAGYQWVEDVL-----ATTAQSAF-----TEFEGLAETRAVEA- 3317
QY 411 SNDLIATRPYN-WNNTIVEPPEQIOND-QVSEQSSSSTPEAQVDESLIEVETVE 468
DB 3318 ----AATYINEMADAVL--DLKQIQNGVELLRKREKQRTVSOBOESKAIVPTAA 3370
QY 469 LTQGIILMDISPIEFSASNLIDOKLNLVAAKARHLYDMPDRVLAINHDDGVN-----RS 522
DB 3371 APEPQPIQEPKMTGPHL-----LYE-----IOSEDDFTYSSSITE 3408
QY 523 ILGRISDAVASAVARAILPDESENEVIDLP-----RTALANKRTPADVYQ----- 567
DB 3409 IWEKVFPAVQVARRA-----HGLTFLPGLADMGIGMIGKTALXYLIQPLGVE 3461
QY 568 --SKVPLY-----VEVADKPRDQIG-----LNGSDPTGTLVLT 601
DB 3462 KCSKYPPIKHKRNGNVSTANGAHGNGLGGSSASALSVSGDSSHLLDYGSDQ-----D 3516

```

OY 602 KREHNL-----NRDYOAGAEELRLSEDKKGYLATKPLSHPLNDOLRATLGYYOQ 652
DB 3517 ELEENAYDCARCEPYSNRSEYDMFMSWLASHRRKQPIQVY-----QPSDNELVPRRG----- 3568
OY 653 EYFGHSTNCFDLSTLTLEHISRSIIIONGWNRTYSLKRYRLDKLTQADPEWMDLPVDF 712
DB 3569 -----TGSNLPMAMKYRTLK-----ETRYKDYGVGF 3593
OY 713 ---VNGK-----PSOEALL--AGVAHVKTVDNLVPMRGYRORY--SLEVG----- 752
DB 3594 RSHINGRGICXCKDIEAGMVEYAGELIRSTLTDK-----RERYDSRGIGCYMFK 3645
OY 753 -SSGLVSDAMN 762
DB 3646 IDDNLVVDATM 3656

RESULT 21
CLPB_HAETN STANDARD; PRT; 856 AA.
AC P44403;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE CLPB protein.
GN CLPB OR H10859.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae.
NCBI_TaxId=727;
RN RA
RP SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
RA Fieischman R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodex A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Fiteback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Gnehm C.L., Fitchman J.L., Fuhrmann J.L., Geochagen N.S.M.,
RA Venter J.C., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT RD.
RL Science 269:496-512(1995).
CC -1- SUBUNIT: THOUGHT TO BE AN APPASE SUBUNIT OF AN INTRACELLULAR
CC ATP-DEPENDENT PROTEASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: U32767; AAC2518.1; -
DR TIGR: H10859; -
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003959; AAA_ATPase-centr.
DR InterPro: IPR001270; Chaprinin_c1pA/B.
DR InterPro: IPR004176; CLP-N.
DR Pfam: PF00004; AAA_2.
DR Pfam: PF02861; CLP-N_2.
DR PRINTS: PR00300; CLP_PROTASEA.
DR PRODOM: PD000739; GSPIL_E_1.
DR SMART: SM00382; AAA_2.
DR PROSITE: PS00870; CLPB_1; 1.
DR PROSITE: PS00871; CLPB_2; 1.
KW Chapterone; ATP-binding; Repeat; Complete proteome.

```

```

FT DOMAIN 161 409 I.
FT DOMAIN 531 722 II.
FT NP_BIND 206 213 ATP (POTENTIAL).
FT NP_BIND 605 612 ATP (POTENTIAL).
SQ SEQUENCE 856 AA; 95837 MW; B8958ED9BD03EA3B CRC64;

Query Match
Best local similarity 19.3%; Pred. No. 7.1; Indels 147; Gaps 25;
Matches 109; Conservative 96; Mismatches 212;

OY 28 ALAQONNPANIINHVDPAHDTAINQAKAGNPVLLTPEQIARLNAAGLNKAPSOALDVV 87
DB 115 ALEERGTISDILKKCGAKKEQISOA-----I0HIRGGQNVNDQNAEERQALEKX 164
OY 88 NPDQSPISRGQSPPLGLDMKVIETTPILSEELFAQESTMGINPDNIPRYGGEOP 147
DB 165 TI-DLTARAESGKLDPVIGRD-----EEI-----RRATQVLRRTK 199
OY 148 NSEVVVPTLEP--EKGILRLYARLFNDGVNKPVR-LKAKPYOSSQSETSATISSHQ 204
DB 200 NNPVLLG---EPGVGKTAIVEGLAQRIVN---GEVEGILKNKRVLSLDMG--ALIGAKY 251
OY 205 KTEPVANIKALEDITQESAMDNGSIPRLRQTALVAARAVGYDIDLSIRNSIGEVY 264
DB 252 RCFEERLKAVALNELSKEGRVIL-FIDEI-HTMVGAKTDGMDAG-NLKRSLARGE- 307
OY 265 ITHDGEPIYIDYRAVEVGEADGAKF-----TTVADEVPLIG-----DVFHGKY 312
DB 308 -IHCAGATTLDEYRT-IEKDAALERFQKVFVDEPSVEDTIALINKERYEIHINVDI 365
OY 313 ETKNLIENASAEHGFYDGRWLDRSYVILPNTADVSLIYDTGTQYRDEVVFTIDPK 372
DB 366 -TDPALVAATLISHRYISRQ-----LPDKALD-----IDDA 397
OY 373 TNQLTTPDPKLPYKELLEQLITVNMGEAVNLQAVFALSNDLIARFYFMKVTEIPEPR 432
DB 398 ASSIMEIDSKPEPLRRLERRITQLEQALOK-----EEDBASRRLEMLKELEKER 453
OY 433 EGIQNDQY-----SFEOSSSRTEPAQVDES-----TLEPYI 464
DB 454 EYAELEEVWKSKAATLSSGQHIKQELDRTKTELEQARRRGDIKAKSELOYGRIPDLKOL 513
OY 465 ETVELTDGILM-----DISPIEFSASNLIDKLLVAKARHLIYMDPDRLVAIN 514
DB 514 EQAERSSEKEMTLNRYVTDDEIAVLASKATGIPYSKM-----MGGEKELR 561
OY 515 HDDGVNRSILGRISDAVAVARAI 538
DB 562 MEDELKRRVIGQ-ERAVDAVANAI 584

RESULT 22
YD86_SCHPO STANDARD; PRT; 1957 AA.
AC Q10411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein Clf3.06c in chromosome I.
GN SPAC13.06c.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxId=4896;
RN RA
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sounos J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

```

RA Gentes S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel K., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Diano S., Gloux S., Lelaure V., Mottier S.,
 RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas R., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revelante J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.,
 RA "The genome sequence of Schizosaccharomyces pombe".
 RL Nature 415:871-880(2002).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z70690; CAA94624.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA0671D9DA CRC64;
 Query Match 2.8%; Score 132.5; DB 1; Length 1957;
 Best Local Similarity 18.1%; Pred. No. 25;
 Matches 168; Conservative 149; Mismatches 322; Indels 289; Gaps 39;
 QY 17 LAAYPLMTSQLAQQNNPANIINHVA-----HDTAINQAKAGNPVLLTPDQIOA 68
 DB 764 LTAANVETTLQDSKAMQSTSLVNSQSTSLNLYHELRDHWNNQSONN-----TLSES 818
 QY 69 RLNAAGLNAKPSOAL-DVY-----NPDQSPISRIQSGSPPLGLDMSVIEETPLSLE 121
 DB 819 KLKTDCEMLTQQNMFLIDWQKLMKHVNOESKVSLEKENVCKSLDKNL----- 869
 QY 122 ELFAQESTEMGINPDYI-----PEYQGEQNSEVYVPTLEPEKGL---I 165
 DB 870 -----RSSLNAVAISDNDQILITQLAEISKNYDSLEQESAQINSLG---KSLAEAKOLHTEN 922
 QY 166 KRLVARELNDGVNKP-RLKAKFYOSOGSETSAIGSHQKTEPYANIKFALEDITQESA 224
 DB 923 EELHRL-----DKLTGKAKIESSKSDGKLT-----ARBEISNLKE--ENMSOSA 970
 QY 225 MDLNSIRLRKOTLVAARANGYIDISTINNSIGEDVY-----ITHDL---G 270
 DB 971 ITSVKS--KLDEITLSSK--LEADIEHLKNKVEVEERNALLASNERIMDLKNG 1024
 QY 271 EPPYIDYRAVEVREGGAD--KATTVADVP--LLIGYVHHGKYEKKNLINASAEH 326
 DB 1025 ENIASLQETIEKKRAENDLOSLSVSEYENLLISS-----QTMSLEDPTN--- 1074
 QY 327 GYFDGRMLDRSVYILPN-----TADVSLYPTGTGYRDEYVFFITDPKTNL 376
 DB 1075 ---QKYTEKNVQKLDKDRQANVELELTSTKYGKLGEBNAOIKELLALRKSSKQHD 1131
 QY 377 TTD-PDKLPVRELLEQLITVNMGEAYNLQAVRALSNDLIATRYNMVTEITVFEREQI 435
 DB 1132 CANFVDDLKEKSDALEQLTNE-----KNELIVLEQSSNSNNEALVVERSL 1177
 QY 436 QNDQVFEQSSSSRTPEAPQVDESTLEPIETVELTDGLIMDISPIEFSSNLQDK----- 491

DB 1178 ANRLSDMKKSLSDSDNIVSIRSDLVRYNDEL---DTLKDKDLSLSTQYSEVCODRDL 1234
 QY 492 -----LNLVAKARHL-----YMPDRVLAIN----- 514
 DB 1235 DSLKGESEFNKAVSLRELCTKSEIDVPVSEIILDDNFVFNAGFSELRLTVLSLENYL 1294
 QY 515 -----HDDGVNRSILGRISDAVSAVARA 537
 DB 1295 DAFNQVFNKMEIILNRLTTDAEFTKVVADEKLQHEHDMWLIQR--GDLEKALKDSEKN 1352
 QY 538 ILPDESE-NEVIDLPETALANRTPADYQSKVPLVYASRKPQDQIGLWGSQDTG 596
 DB 1353 FLRKAEMTENSILSEEGKEETKEIAEL-----SSRLDENQIAT----- 1392
 QY 597 TRLVTKREHNLINDGYQAGAEIRLSD-----KKGVLVYATKP 635
 DB 1393 NKLKNQDTH--LND-----EILKEDVLKEKSLIISLEESLSNOROKESSILDKNE 1443
 QY 636 LSHPLNDQLRATLGYQOEVEGHSTN---GFDLSTRTLE-----HEISRIIIONGWN 684
 DB 1444 LEHMLDTSRKNSSIMKISINSIILDDKSPELASAVKIGALOKLSESLSMEN----- 1499
 QY 685 RTYSLRRLD--KLKIQAPPEIWMOLPVDVNGKPSOE-----ALLAGVAHKTVADN 735
 DB 1500 -----IKSQLQEAKEKIQVDESTIOELDHEITASKNNYEGKLNKDSTIIRLSENIQLNN 1555
 QY 736 LVNPMRGYRQRYSLVSGSSSLVSDANMA 763
 DB 1556 LLAERSAVRLSTKESEIILQFNSRLA 1583

RESULT 23
 HS75_KLUMA STANDARD; PRT; 612 AA.
 AC P41770;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Heat shock protein SSB.
 GN SSB.
 OS Kluveromyces marxianus (Yeast) (Kluveromyces fragilis).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluveromyces.
 NC NCB1_TaxID=4911;
 RN [1]
 RC SEQUENCE FROM N.A.
 RA STRAIN=ATCC 12424;
 RL Iborra F.;
 RL Submitted (MAY-1991) to the EMBL/Genbank/DDA databases.
 CC -I- FUNCTION: MAY AID IN THE PASSAGE OF THE NASCENT POLYPEPTIDE CHAIN
 CC THROUGH THE RIBOSOME CHANNEL INTO THE CYTOSOL. SUCH AN INTERACTION
 CC COULD BE CRUCIAL FOR CONTINUOUS TRANSPORT OF THE POLYPEPTIDE.
 CC COULD SERVE TO PREVENT THE NASCENT POLYPEPTIDE FROM INTERFERING
 CC WITH TRANSLATION BY CLOGGING THE RIBOSOME CHANNEL (BY SIMILARITY).
 CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH TRANSLATING
 CC RIBOSOMES. MAY BIND DIRECTLY TO THE NASCENT POLYPEPTIDE.
 CC -I- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X59963; CAA42589.1; -
 DR HSP: P08107; IHQO.
 DR InterPro: IPR001023; HSP70.
 DR Pfam: PF00012; HSP70; 1.
 DR PRINTS: PR00301; HEATSHOCK70.
 DR ProDom: PD000089; HSP70; 1.
 DR PROSITE: PS00297; HSP70_1; 1.


```

Db 544 KHTKYBOLDEYRSMHLEDLRQKTYDMNNMAMAVSNIGDIE-----ALRKQKDEYSKLK 598
OY 501 HLYDMPDVRVLAINHND-----DGVNRSILGRISDAVSAVARAI-LPDSENVIDLPERTA 555
Db 599 D-----AEDRKHEIESEFPDINSYPTSTIGKTEDVRLLEPOIKLAEDLKROREPLREKVK 654
OY 556 -LANRKTADYOSKVPVLYFVASDKPR-----DGOIGLWGSDDTGLRYLTFEHLINR 610
Db 655 DLNRSRSGMDEIOKRKNELSVKASESETRLKYVEGOIQAATLSLSGKR--SKVE-----TL 708
OY 611 DGQAQAGELRLSEDKKVKYLAATKPLSHPLNDQLRATLIGVQOEVFGSTNGFDLSTTLE 670
Db 709 RSHVSETEORISDREDE--RMKKIEKAINDVKRI-----REAFGR-----748
OY 671 HEISRSIIQGNMNRYSRLRYRLDKLTKQAPEPTWQDLPDVFVNGKPSQALLAGVAVHK 730
Db 749 -----NGVFAM-----IRO 757
OY 731 TVADNLVPMRGYRORYSLEVSGSLVSDANMAIARAGI 769
Db 758 SVSDYLAKTRDYLSFDFDIDISVDQDENNVYRGV 796

```

RESULT 25

RPOB_UREPA

ID RPOB_UREPA STANDARD; PRT; 1434 AA.

AC 09PQV6;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6) (Transcriptase

beta chain) (RNA polymerase beta subunit).

GN RPOB OR U0187.

DE Ureaplasma parvum (Ureaplasma urealyticum biotype 1).

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.

OX NCBI_TaId-134821;

RN (1)

RC SEQUENCE FROM N.A.

RC STRAIN-Serovar 3;

RX MEDLINE-20500219; PubMed-11048724;

RA Glass J.I., Letkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,

RT Cassell G.H.;

RT "The complete sequence of the mucosal pathogen Ureaplasma

urealyticum.";

RT Nature 407:757-762(2000).

CC -I- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION

CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS

CC SUBSTRATES (By similarity).

CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +

CC (RNA)(N).

CC -I- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE

CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1

CC BETA' CHAIN (By similarity).

CC -I- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@sib-sib.ch).

CC EMBL: AE002118; AAF30594.1; -

DR HSRP: O9KWU7; 1HOW.

DR InterPro: IPR001572; RNA_pol_B.

DR Pfam: PF00562; RNA_pol_B; 1.

DR ProSite: PS01166; RNA_POL_BETA; 1.

KW Transferase; Transcription; DNA-directed RNA polymerase;

KW Complete proteome.

SQ SEQUENCE 1434 AA: 162022 MW: 2841BFA50C19BA7A CRC64;

Query Match 2.8%; Score 131; DB 1; Length 1434;
 Best Local Similarity 19.2%; Pred. No. 19;
 Matches 158; Conservative 132; Mismatches 281; Indels 254; Gaps 43;

```

OY 56 NPPVLLIPEQIQARLNAAGLNAK--PQSQALDVNFPDDQSPISRIEQSPPILGDMSTVEE 114
Db 641 NPLAELTNKRRISSAMGPGGISREDPNLDIRDY-----HSHSSSRICPIETPGMNIIGLI-- 694
OY 115 TTPLLELFAQES-----TEMGINPNDIPEYQSGQPSSEVVVPT---LEPEK 161
Db 695 ---SLSLAKVDENGFIIVAPYVEEDGVKEDY--KFLAHEDNNTYIAESSVQLEDNK 749
OY 162 PGLIKRLYAR-----LFN-----DGVNKKVPLKAKFYOS-----SQSGEFAIGSSHOK 205
Db 750 RILDEQVYARRGSTGLFSPREVDIFDIYPRQVYSIASAIPFIENDDGAALMGSMNOR 809
OY 206 -----TEPYANIKALAD--ITQESAMDLSNGISPRLOTALVAA--AVGYDIDLSTIR 256
Db 810 QAPPLIKPYAPIVGTGTETFKIAHDSGM-----AVAKNDGVLEFVDSQKIIIR 857
OY 257 NSIGEVDVTHDGEPIYIDYRAVEVGEADKAFITTVADVEPL-----IGD 305
Db 858 NDNDKLD-----DKLIKIRKSNOD-----TCNNQIPYVAGORVHKSETIGD 900
OY 306 --VFHGGKYETKKNLIENASAEHGY-----FDGRMLDRSVYIL-----342
Db 901 GPANQNGELALGRNIIIVGYTTRRGYNFEDAIISEKRLVDQDFSHIHDEITQCMKTN 960
OY 343 --PDNTRADVSLIYDT-----GTQYRDEVVFTIDPKTQLTTPDKL----383
Db 961 GDEEITRDMNPNVSDPTAKRFLDNGIVLVGAEVHGDVAVGKTPRGVNETAPEDRLQTI 1020
OY 384 --PYKRELLEQLLVNNGEAVNLQAVR-----ALSDLIATRYFMNVNTEIYFPE 431
Db 1021 FGDKSKTVKDSKLKVGQGIYAAVARIKSSDENGSELPDVI---ELIKYIV--Q 1073
OY 432 REQIQ-NDQVFEQSSSSRTEPAQVDESTLEPYETVELTDLGILMDIS-----PIEESA 484
Db 1074 KRRIQVGDMMKAGRGNGKIY-----SKVYPIODMFLKDGTPLDLMLNPLGPSKMI 1126
OY 485 SNLIQDKLNLVAAK--ARHLYDMPDVRVLAINHND-----GVNRSILGRISDAVSAVARA 537
Db 1127 GQILELHLGYAAAEIGKQLQIAIDQ--LGYEKYISLFGINELIAKKIYENISNLIKH 1183
OY 538 ILPDESENEVIDPERALANRKTADYOSKVPVLYFVASDKPRQOIGLGSGSTGT 597
Db 1184 KQAKQAKD--IDLIDVTYILK-----ELGLSY-DDIGI 1213
OY 598 RLVTYKF-----EHNLI-----NRDGYOAGELRLSEDKG--VKLYATKPLSH 638
Db 1214 KISTPVVDGANHDDIYSINMEANIDENKKGQVLDGRKTGEPPDGLISGLTYMLKLDH 1273
OY 639 PLNDQLRA-TLG-----YQGEVFGSTNGFDLSTRLLEHISRSIIQNGMN--RTYSRLY 691
Db 1274 MVDDKIHRSVGPYSKITQGPLGKSGONG--GQREGMEV-----WALEAIGAAV 1321
OY 692 RLDKLTQAPEPTWQDLPDVFVNGKPSQALLAGVAVHKVADNL 736
Db 1322 NLEELITL-----IKSDVOGR--NQAYNAIINGHDVAVADGM 1355

```

RESULT 26

SCA4_RICEP

ID SCA4_RICEP STANDARD; PRT; 981 AA.

AC 09A37;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Antigenic heat-stable 120 kDa protein (Pst120) (120 kDa antigen)

DE (protein ps 120) (Fragment).

GN SCA4 OR D.

OS Rickettsia felis (Rickettsia azadi).

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

CC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=42862;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sekegova Z., Roux V., Raoult D.;
 RT "Phylogenetic analysis of Rickettsia spp. by comparing the sequence of
 RL gene D coding for an intracytoplasmic protein."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AF196973; AAK31304.1; .
 KW Antigen.
 FT NON_TER
 FT TER 981
 SQ SEQUENCE 981 AA: 107514 MW: 7F18F421E2C262E1 CRC64;
 Query Match 2.88; Score 130; DB 1; Length 981;
 Best Local Similarity 18.4%; Pred. No. 12;
 Matches 166; Conservative 136; Mismatches 348; Indels 252; Gaps 39;
 OY 62 TPEDIOARLNAGLNAKQSOALDVNFDDOS-PISRIGESPP-GLDSVI-----EE 114
 DB 23 TPQELAA---DDGFSFPASSSTQSPISISSGSISSDSQSDPTTKAVRETIIOPOKDE 79
 OY 115 TPPLSLEELFAQESTEMGINDYIPEYQGEQNPSEVVVPTLEBEPGLIKRIYA--- 170
 DB 80 IAEIOLKALADLRDL-----AEQKRKEI-----EEEKDITSAFGNPAN 121
 OY 171 RLFDNGVAVKRLAKFQSSOSGCTSAIGSSHOKTEPYANIKALEITQESANDLNGS 230
 DB 122 REFIDKALENELKLLKLEISIAIGKNVL-STYSAANGYQGFQVQENQISASDLNAT 180
 OY 231 IPR-----LROTAL-----VAARAVGYVDLSTIIRNSIGEVDYIIDLGPYID 276
 DB 181 VVRNDAGDELCTLMTETVYKTKPFYAKQDG-----TQVQINSYRIDP-----PIKLD 228
 OY 277 -----YRAVEVGEG---ADKA---FTYVADEYPLLIGVFHHGKETKKNLIENASA 324
 DB 229 KADGSMLSMVALKADGPKPKDKAVFTAHYEBGP-----NGKPOLK---EISSP 276
 OY 325 EHGFFDGRMLDRSYDVLIPONTADVSLYDTGTQYRFDEVEVFTIDPKTNOLTPDPKLP 384
 DB 277 KPLKFAG-----DGPDAVATIEHGEYI-----TLA 302
 OY 385 VKRELLLEQL---TVNNGEAVNLAVRALNSDLIATRYFNWNTVEYPEREOIONDOVS 441
 DB 303 VTRCKYKEMREVELNQGOSVDLS--OTIADLTQVQGRSOETPOPIITTPQDELKS---S 357
 OY 442 FEQSSSRTPEAQVDESLPEVIEFTVETDGLIMDISIEFSASNLIDOKNLNVAAKARH 501
 DB 358 IETPTTVOVPIITPANOPLOPETSQMPQOVNPNLLYATSLSTMODLLNVA--- 413
 OY 502 LYDMPDDRVLAINHDDGVNRSILGRISDAVASAVARAILPDESEN-----EVIDLPERTA 555
 DB 414 -----CLTEKQDNKOI-----DLNEAATAILNNKSDIAEKQANITALENTV 458
 OY 556 LANRKP-----ADYQSKRVPLVEVA-----SDKPRDCQIGLQW 591
 DB 459 NNNDLTPTDVKAGVNAVLETIKNDQNPDPDESKSMLEATVAIALNSEQEKQKQOMLEK 518
 OY 592 GSDVGTGLVYKFEENILNR-DGY-QAGAEALRLSDKKGVKYLAKRPLSHPLNDQLRALG 649
 DB 519 AVDVGSL--KDDASRYVAIDGITDAVAKSNLSTEDKQTMILAV----- 560
 OY 650 YQDEVFGHSTNGFDLSTRTLEHISRSIIONGNRTYS-----LRYRLDKLTQAPPE 703

DB 561 -----CDKVASELSNAE-KOKLLGSVLKKGVAEAOVLSPEQOOLMOQNLDKI-TAEQTK 612
 OY 704 TWQDLPVDFVNGKPSQDELLAGVANHKTVADNLVPMRGYGRVSLVEGSSGLVSDAMA 763
 DB 613 NAQITVEYGILANPAFNTIAKTEALQNTVTKVLSDFIAETKGETLE-----SITKV 664
 OY 764 IARAGIS-----YSPGDNAYGSRNRAHOMTGIGIAGYIWSDNFNHVPRLRFPAGDQ 817
 DB 665 VAESPLNGQDKADIVKNGE-ALASHKTMAPTEKIST-----IE 702
 OY 818 SIRGAHDSLSPSIDSKGYLTGGQVLAAGTAENYEFMKDLRAVFGDIGNAYDKGFTMDT 877
 DB 703 SVEKQVASEITDLEBKRLMTKGLVGIEGKANPEITSEKTKAV-----SRGIDKST 754
 OY 878 KI 879
 DB 755 AI 756
 RESULT 27
 PGCV_RAT STANDARD: PRT: 2738 AA.
 ID PGCV_RAT
 AC QPERBA: 008592; 088564; QPR1K4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Versican core protein precursor (large fibroblast proteoglycan)
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glia
 DE hyaluronate-binding protein) (GHAP) (Fragments).
 GN CPBG2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0), SEQUENCE FROM N.A.
 RP (ISOFORM V3), AND SEQUENCE OF 2657-2738 FROM N.A. (ISOFORM VINT).
 RC STRAIN=Wistar Kyoto;
 RX MEDLINE=99327053; PubMed=10397680;
 RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
 RA Wright T.N.;
 RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
 RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
 RN [2]
 RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0).
 RC STRAIN=Wistar Kyoto;
 RX MEDLINE=98308094; PubMed=9642104;
 RA Miley P., Maurel P., Chiba A., Mevissen M., Popp S., Yamaguchi Y.,
 RA Margolis R.K., Margolis R.U.;
 RT "Differential regulation of expression of hyaluronan-binding
 RT proteoglycans in developing brain: aggrecan, versican, neurocan, and
 RT brevican.";
 RL Biochem. Biophys. Res. Commun. 247:207-212(1998).
 RN [3]
 RP SEQUENCE OF 2421-2463 FROM N.A. (ISOFORM V0).
 RC TISSUE=Kidney;
 RX MEDLINE=98094159; PubMed=9434070;
 RA Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
 RT "Proteoglycan expression in the normal rat kidney.";
 RL Nephron 77:461-470(1997).
 RN [4]
 RP SEQUENCE OF 2535-2738 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Lung;
 RA Blomberg L.A., Chan W.-Y., Clerch L., Massaro D.;
 RT "Molecular cloning and characterization of two developmentally
 RT regulated genes in rat lung.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.

[illegible]

QY 463 VIET-----VELTDSILMDISPIEF-----SASNLIDQKLNLAJAKA---RHLYDMPDRV 510
Db 1230 SVETSSSSPTIGLAPSTVDLIEVEYMNQTSKTLISLSEKPTSOAEVRDLY----- 1282
QY 511 LAINHDDGVNRSLIGRISDAVSAVARAILPDE-----SENEVIDPERTALANRKTAPD 564
Db 1283 -----PGIGEDPFGSGDSE-YPTVSTTKMEELVGMGSENERV-----KD 1321
QY 565 VYQSKKVPILYVYASDKPRDGOIGLWGSDDTGRVLTKEFHNLINRQYQACAEPLRSED 624
Db 1322 TQTLSSIP-----PTSDNINPVPDSKGFSTVASTAFPMWEEFTSAEG--SGEE--LSSV 1373
QY 625 KKGVKLYATKPL-----SHPLNDQLRATLGYQOEVEFGHSTNGRDLSTRLEHEISRS 676
Db 1374 RRSVSL--VPLGLVDLPTTESPFQDEFEAAAVTAEAKQSAALPLAVSGNTVDLTENRD 1431
QY 677 ITONGMNRFTSLRYRLDKLKTQAPPEY-----QDLPVDFVNGKPSQOAL 722
Db 1432 IEVNS-----TMS-----VDLPQTEPAKLMSKPEVNEPEKQEGSETVTQDKAQGKSFESL 1483
QY 723 LAGVAVHKTAVD-----NLVPMRGYRQRTSLEKSGGLYSDANMALIARAG 768
Db 1484 HSLALPEQTLTSSQSLIETEVOTSYSLMTTKYTNFNEVEEGETSI---AHMSTPGCQ 1540
QY 769 ISGVYSGDNAYGSRRAHOMT 789
Db 1541 IKGLESYPTHPEATGKSYSPS 1561

RESULT 28
SLAP_BACCI STANDARD: PRT: 1616 AA.
AC P35824:
DT 01-JUN-1994 (Rel. 29. Created)
DT 01-JUN-1994 (Rel. 29. Last sequence update)
DT 01-FEB-1996 (Rel. 33. Last annotation update)
DE S-layer related protein precursor.
GN BUTB.
OS Bacillus circulans.
OC Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1397;
RM [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B3312;
RX MEDLINE=94374689; PubMed=7522196;
RA Aubert-Pleyert E., Davies J.;
RT "Biosynthesis of butirosin in Bacillus circulans NRRL B3312:
RT Identification by sequence analysis and insertional mutagenesis of
RT the butB gene involved in antibiotic production."
CC - FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. MAY PLAY A ROLE
CC IN THE EXPORT OF BUTIROSID FROM THE ORGANISM.
CC - SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L20421; AAA62588.1;
DR InterPro: IPR003343; Big_2.
DR InterPro: IPR001119; SLH.
DR Pfam: PF00395; SLH; 1.
DR Pfam: PF02368; Big_2; 1.
KW Signal; Cell wall; S-layer.
FT SIGNAL 1 30
FT CHAIN 31 1616 POTENTIAL.
FT LAYERIN 1616 S-LAYER RELATED PROTEIN.
SEQUENCE 1616 AA; 172874 MW; 1F03FF7A4F113AA7 CRC64;

Query Match 2.7%; Score 129.5; DB 1; Length 1616;
Best Local Similarity 17.5%; Pred. No. 27;
Matches 203; Conservative 169; Mismatches 407; Indels 379; Gaps 51;

QY 34 NPAINIHVPAHDTALINQAK-----AGNPPLPPEQIQARLNA----- 72
Db 189 NPDPGLNLVTASCTTYTSPFTYSGOVVRKSGQGISLTYNTEALAKADGAFQSAVRLT 248
QY 73 AGLN-----KPSQALDVNEDDPSISRIGESPPGLMDSVTEFTPLS-LE 121
Db 249 AGTNALSVKLLKGRREIVSTVTYVNDAAQPADLLEVAAPIDITISGPAHAIGYVD 308
QY 122 ELFAQESTENGIPNDIYPIEYQGEQPNSEY-----VPEPTLEPEKPKILKILYAR 171
Db 309 QDIAGIDDTVALEFTNDWGPQITVPQFNVAAYQVDAGSKVTKVNPSPIDGKTPAWTGPTDLE 368
QY 172 LFNQGVNKKVPRLLKAKFYQSSQSGSETSAIGSSHOK--TEPYANIKAALE-----DITQESA 224
Db 369 IPSGGYVLAQ-----DTSYAGKNIKKIYATYKVGDAIKRKNGFAVVPYKDL 416
QY 225 MDLNGSIPR--LRQALVAARAVGYDIDLSTIRNSIGEVDYIIDLGPVYIDYRAVEY 282
Db 417 MGTGCPYIARVTLNDNYAM-----YTETKPSY-----ELSGITIMDDPSKI---ALTY 460
QY 283 RGE-----GADDKAFT--TYADEVPLLIGDFHGHKYEKTKNLEMASAEHGYEDGR--- 332
Db 461 NGTLPFPFGDCKFKTSTYTLAEGINTDLYVTKGKGRDQSKDLV--YSRGFSSTGAKVYL 518
QY 333 WLDRSVDVILPONTADVSLIYDTGTQYREDEVVEFT-----IDPKTNQTLTDP---DK 382
Db 519 WVDQANARKFQGTGDVNAVNFRLPAKEKNGVTSVVEYKGVGYSYKSTLTGRPYSAIK 578
QY 383 LPYKR-----ELLEQL-----LTVMGEAIVLQAVRALSNLLA-----TRYEN 421
Db 579 APEKASNDPLDLQEFIRYSRELGDIVHSFNIFEGSIAISNEFALLDSHLMDEERYN 638
QY 422 MYVTEIYPEREIOINDOVSFEQSSSRTEPAQVDSLEPVEYELTGLIMDSPIE 481
Db 639 AADNGCIKRLRESAKGANAFAVNPSPNDEVRDFOL--KTIEEVLQANDV--DGVLLDARD 695
QY 482 FSA---SNLIDQKL-NLVAKARHLYDMPDRVLAINHDDGVR----- 521
Db 696 NESADFSDLTKAKFESFLARGKOLQNMPPD--VFYAGNVAKDGPLLRDMWEPFSKTI 752
QY 522 ---SLIGRISDAVSAV-----ARAILPDES- 543
Db 753 KSFTSEVRQLTDRVKAKEGKKILEVSAVYGSWFESYLVNGVHMGSTEFRRDERLRMDKSV 812
QY 544 -----ENEVIDLPERTA--LANRKTAPDVYQSKKVPILYVYASDKPRDGOI 587
Db 813 YTPGYEESGVYKMLDITMIGAVQTTAPLEHYITLONITNGVPLXAGA----- 863
QY 588 GLWGSDDTGRVLTKEFHNLINRQYQACAEPLRSEDKKGKLYATKPLSHPL----- 640
Db 864 -----LTNVQEPALQRFVQAGL-----VNTGILMIFDASQVWMPAGALRN 906
QY 641 -----NQQLRATLGYQOEVEFGHSTNGRDLSTRLEHEI 673
Db 907 LYYVRDYQIGISLPDPSDFLBSGIYNTNLIENNIGVLTDFGTSTGNSRFGV----- 959
QY 674 SRSIIIONGM--NRTYSLRYRLDKLKTQAPPEYMODPL-----VD----- 711
Db 960 -EAYVDSSGKVTSPKTKQAMTMNMCK-----PDEINSVIIPKGFVYVSLDASGIRTKRQ 1013
QY 712 -----FVNGKPSQOALLAGVAVH--KTYADNLVPMRGYRQRTSLEKSGGLVS----- 758
Db 1014 LVANAAYETGDSYRAAALSGFLAYEGLRTSADSVT--FRG-----KVDYLGPGKASVTYNG 1066
QY 759 -----DANMALIARAGISG-----YVSRGDAV-----YG 781
Db 1067 QERALREDGTFOADVITRPGANPVYIIVRVGDKTNEKTVTTIIGDEAAVAKALKLDRGYS 1126

QY 782 SNRAHQTGCIQAGYINSNDNHNHVPYRLRFPAGDQSTIRGY-AHDSLSPISDKGYLTGQ 840
Db 1127 MNKESLRLVATAYE-----SSSSTDVYTGQAAYASLDPAVVSVDAT-GR 1169
QY 841 VLAV----GTAENYVE-FMKDLRLAV-FGDIGNAYDKGFTMDTKIGAG-VGVRRASPYGQ 893
Db 1170 ITALRESSGVQATYBEHTATATARSVTSGSTGGSDTGGTGGSGGAGGGGTARSGP 1229
QY 894 VRVDVATGVKEGNPIKL 911
Db 1230 ERTSVETKDSGRNLT 1247

RESULT 29
FRPA_NEIMC
ID FRPA_NEIMC STANDARD: PRT; 1115 AA.
AC P55126;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN Iron-regulated protein frpa.
OS Neisseria meningitidis (serogroup C).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=135720;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-FAM20 / Serogroup C;
RX MEDLINE-93139051; PubMed-8423153;
RA Thompson S.A., Wang L.L., West A., Sparling P.F.;
RT "Neisseria meningitidis produces iron-regulated proteins related to
the RTX family of exoproteins.";
RL J. Bacteriol. 175:811-818(1993).
CC -I- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL
DISEASE.
CC -I- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED.
CC -I- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN SECRETED
CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
ACTIVITY.
CC -I- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L06302; AAA25454.1;
DR InterPro: IPR001343; Hemlysn_Ca_bind.
DR Pfam: PF00353; hemolysincbind; 7.
DR PRINTS: PR00313; CABINDNGRPT.
DR PROSITE: PS00330; HEMOLYSIN_CALCIUM; 5.
KW Toxin; Calcium; Outer membrane; Repeat.
FT DOMAIN 765 988 13 X REPEATS, GLY-RICH.
FT REPEAT 774 779 2.
FT REPEAT 783 788 3.
FT REPEAT 902 907 4.
FT REPEAT 911 916 5.
FT REPEAT 920 925 6.
FT REPEAT 929 934 7.
FT REPEAT 938 943 8.
FT REPEAT 947 952 9.
FT REPEAT 956 961 10.
FT REPEAT 965 970 11.
FT REPEAT 974 979 12.
FT REPEAT 983 988 13.
SQ SEQUENCE 1115 AA; 122177 MW; 21841065BCCACEFA CRC64;

Query Match 2.7%; Score 129; DB 1; Length 1115;
Best Local Similarity 19.9%; Pred. No. 17;
Matches 205; Conservative 102; Mismatches 335; Indels 388; Gaps 45;

QY 56 NPVLLTPEQIQARLMAAGLNAKPOSQALDVYVNDQSPISRGQSPPLDMSVIEET 115
Db 155 NSRMKLSSTVYQHHEINLYGFKIK-----DYN-----ERLGE-----SIMIDDF 194
QY 116 TPLSLLELFAQESTEMGINPDYIPEYQGEQPNSEVVPPLLEPERK 163
Db 195 TPKSIANFEA-----DPTYSNVLEEVSRFTYSLVPPDANPWKGGEDYIGRGISEWG 246
QY 164 -LIRLYARLEFNDGVNKNVPRKAEYQ-----SSQGETSAIGSSHQKEPYANIKAALED 218
Db 247 ELLEKWKQDF-----LPYLEKEMDQPKFREDMLPERPEARERMKIDPRKSGKINHYD 300
QY 219 ITQESAMDNGSIPRLKOTALVAAR--AVGYIDIDLSIIRNSICEVD---VIHIDGEP 272
Db 301 ---PLALDLDGD-----GIETVAAGFAGALFDRHNOGIRFATGVASADDLVRLDNGN 352
QY 273 VYIDYRAVEVREGADDKAFTTVADEVPLLIGVFHHGKYTEKKNLENSAEHGYFDGR 332
Db 353 GIID-----NGAE-----LEGD-----NTR--LADGSFAKHGYALA 382
QY 333 WDRSVDVILPDNTADVSL-----IYDTGTQYRFDEVEVFTID----- 370
Db 383 ELDSNGDNIT--NAADAAPQLRWQDLDNDGISQANBELRTLELGIQSDLAIKDYKNK 440
QY 371 -----PKTNQLTTD-----PKLIPVRELLEQLLVNMG EAY 402
Db 441 LGNGNTLAQOGSYTKTGTAKMGDLLAADNLHSRFKDKVELTAEDQAKANLAGIRLR 500
QY 403 NLAOVRALSNDLLATRFENNVNTEIVEPEREQIONDVSFEQSSSTEPRAVDSESTLEP 462
Db 501 DLREAAALSGDLA-----NML-----KASAAETKEAQL--ALLDN 534
QY 463 VIETVELTDGIILNDISPIEESA-----SNLIODKLNLYAA--KARHLYDM 505
Db 535 LIHKMAETDSNMCKKSPMRSTDMTOTANEGIALTPSQVQALKNALYSLSDKKAIDA 594
QY 506 PDDRVLAINHDDGVNRSILGRISDAVASAVARAILPDESENEVIDLPRTALANKRPADY 565
Db 595 ARDRIAVLDAYTGQDSSSTLYMS-----BEDALNIVKYVNDT 631
QY 566 YQSKKVPY---VEVASDKRRDQIGIGMSDPTGT---RLVTKFEH-----NLINR 610
Db 632 YDLAKNITONLLFYRLQPLYNQISFKMENDFITDFSGLYOAFNHVKEPNQKAFVDL 691
QY 611 DGYQAGAEIR-----LSEDK----- 625
Db 692 AEMLAAGELRSWYEGRLMADYVEEAKKAGKEDYQKVLQETVALIAKTSGTQADDILQ 751
QY 626 -----KGKLYATRP---LSHPLNDQLRATLYGOQEVFGHS----- 658
Db 752 NVGFGHNKNVSLYGNNDGNDTLIGAGNDYLEGSSGSDTYFGKGFQDPTYYNNDYATGKR 811
QY 659 -----INGF--DLSTLTLE--HEISR-----SIIONGMMRTYSLARYDLKLT 698
Db 812 DIIRFTDGTIAMLTLFTREGNHLIDIAKDDSGQVTVQSYFQNDG-----SGAARIDE--- 863
QY 699 QAPRETWQDLPRVDFVNGKQSQSEALLAGVAUHKTVADNLNPNMNGYRQRYSLVGS--SGI 756
Db 864 -----IHDNGKLVDA-----TVKELVQOSTDGSRLVAYOGSGLTNG 903
QY 757 VSDANMALIARAGISGYVSFGDNAYGSNRAHOMTGGIOAGYINSNDNHNHVPYRLRFPAGD 816
Db 904 LGD-----DLYLGADGNDLLNGAGNDSIYSGN-----GND 934
QY 817 QDIRGYAHDSLSPISDKGYLTGGVYLAVTAGYNTPEFMKDLRLAVFGDIGNAY-DKGFYN 875
Db 935 TLDDGEGNDALTYGNGDNLNGE-----GNDHINDEGND--TLIGAGNDYLEGSSGS 987
QY 876 DTKI-GAGVG 884
Db 988 DTYVEGEGRG 997


```
Db 762 SNKPLEDLEPEKOVEVAKSLSEKTSVATYA-----GGAVVGATKTHSATKASTA 812
Qy 631 YATKPLSHPLNDOLRATLGIQOEYFGHSTN-----GFDSTRT-----LEHE 672
Db 813 ASKSKSTETL--VMKTTASTSTGYGANKSAAPRPARLGIKSTATITSTTSLTGN 870
Qy 673 ISSTIIONGGMNRTYSLRYLDKLKTQAPPEWODLPVDFVNGKPSOELLAGVAHRTV 732
Db 871 PRKSLSSMNG--STVKPPTLSTGRPATAPVSKVTLGAKITTKNPTASG-----TA 919
Qy 733 ADL-----VPMRGYRQRYSLFVGSSGLVSDANMAIRACISGYSGDNAYGS 782
Db 920 SDNVTTTTLPLVSTNARRPATSGTGSVAS--STARRPVT-----NAKGS 962

RESULT 31
DPOL_BACSU          STANDARD:          PRT:          880 AA.
AC 034996;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA polymerase I (EC 2.7.7.7) (POL I).
GN POLA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=1423;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Betero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Broillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Erlington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizi A., Galleron N.,
RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisept G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue Y.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Preecean E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche S., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serrot P., Shin B.S., Soldo B.,
RA Sorokin K., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosto V., Uchiyama S., Vandenbol M., Vanlier F., Vassaretli A.,
RA Viati A., Wambuit R., Wedler E., Wedler H., Welternegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -I- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY (BY
```

```
CC SIMILARITY).
CC -I- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -I- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS.
CC -I- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-A FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF008220; AAC00350.1; -.
CC EMBL: Z99118; CAB14869.1; -.
CC HSSP: P52026; 1XWL.
CC Subtilist; BG12656; POLA.
CC Interpro: IPR002562; 3_5_exonuclease.
CC Interpro: IPR002421; 5_3_exonuclease.
CC Interpro: IPR001098; DNA_pol.
CC Interpro: IPR002298; DNA_pol.
CC Interpro: IPR000513; EXO_N_I.
CC Interpro: IPR003583; HHH_1.
CC Interpro: IPR003584; HHH_2.
CC Pfam: PF00476; DNA_pol_A; 1.
CC Pfam: PF01367; 5_3_exonuclease; 1.
CC Pfam: PF02739; 5_3_exonuc_N; 1.
CC PRINTS: PR00868; DNAPOLI.
CC SMART: SM00474; 3SEXOC; 1.
CC SMART: SM00475; 53EXOC; 1.
CC SMART: SM00278; Hhh1; 1.
CC SMART: SM00279; Hhh2; 1.
CC SMART: SM00482; POLAC; 1.
CC TIGRfams: TIGR00593; polA; 1.
CC PROSITE: PS00447; DNA_POLYMERASE_A; 1.
CC Transferrase; DNA-directed DNA polymerase; DNA replication; DNA repair;
CC Hydrolyase; Exonuclease; DNA-binding; Complete proteome.
CC KW SEQUENCE 880 AA; 99091 MW; AFB434ARDF26FA48 CRC64;
CC SQ

Query Match 2.7%; Score 128.5; DB 1; Length 880;
Best Local Similarity 17.3%; Pred. No. 12;
Matches 174; Conservative 170; Mismatches 355; Indels 307; Gaps 46;

Qy 8 ANSEFMVAL-----AAYLPLMTSQLAQGNPANTINHPAHD-----ALNQA 52
Db 16 AYRAFFPLPLISNDKGYHTNAVIGFAMITLKMLEDKPTHTLVAFDGKTTFRGTRKEY 75
Qy 53 KAG---NPVLLTPPEQ---IQARLNAAGLNAKPQSQALDVVNFDDQSPISRIQSPPLG 106
Db 76 KCGNQKTPPEL---SEQMPTRELLDAY-----QISRYELEQYFADDIIGTLAKSAEKDG 127
Qy 107 LDMSVIT---EETPPLSLIEELPAQESTEMGINPND-YIPEYQGEQPNSEVVVPTLPEKP 162
Db 128 FEYKVFSGDDLIQLATDKT-TVAITRKGITDVEFYFPEHKEKYG-----LTPEQI 178
Qy 163 GLIKRLYARLFPNGCVNKNVPLLAKKFYSSQSGERSAIGSSHQKREPRANIKAALEDITOE 222
Db 179 IDMK-----GLMGSSNDIPVPG-----VGEKTAI-----KLKQPDSEVKLESIDEV 223
Qy 223 SAMDLNGSIRLROTALVARAVGYDIDLSIIRNSIGEVDVITHDGEPEYIDYRAVEV 282
Db 224 SGKKLKKELEFPDQALMSK-----ELATIMTD--APIEVSQSLLEY 263
Qy 283 RGEQAD-----DKAFTVADEVPLLIGVFNHGGYETKKNLIENASABHGFYDGRWLD 335
Db 264 QCFNREQVIAIFPDGLGHFTLLER---LGE-----DSAEAQ-----D 297
Qy 336 RSYDVI---LPDNTADV-----SLIYDGTQYRFDPEVVFETIDPTNQLTDPDKLPV 385
Db 298 QSLIEDIVKTVTDVTSILVSPSAFVEQIGDNTHEPILFSTI-----D 341
Qy 386 KRELLBOLLTVNNGEAVNIQAVRALSNDLIATRYFNMVNTEIIVPEREQIQNDQVSFEQS 445
```

```
Db 342 -----VNETGAFYFKDIAVESEVF-----:KWEVENDOKWVF 375
OY 446 SSSRTTBA-----QVDESTLEPIETVELLDGLMDISPIEFSASNLIOOKLVLVAKA 499
Db 376 DSKRAVVALRMQGIETLGAEPDTLLAIIINPG-----NSYDDVASVAADYG 422
OY 500 RHLVMDPDRVLAINHDDGVNRSI-----LGRISDAVASAVARAILPDESENEYDL 550
Db 423 LHITS-SDESV-----YKGKAKRAVPSESDVJSEHIGRKALAIQIS:REKLVQLENNDOJEL 477
OY 551 PERTALANRKTADVYOSKVPYLVFAVASDKPRDGOIGLGSGSTGTRLVYKFEHNLINR 610
Db 478 -----FELEMPPLALI-----LGEMESTG-----VKVVDRIKR 506
OY 611 DGYOGAELRLSEDKKGVKLVATKPLSHPLNDOLRATL-----GYOQVEFGSTNGCFD 663
Db 507 MGEELGAKKLEYEK-----IHEIAGEPFINSPKOLGVLIFERIGLPYVKKITGYSTSDV 564
OY 664 LSTRLEHEISRSIION-----GGMNRTY-----SLRYRLDKLKTQ-----APP 702
Db 565 LEKLADKHDIYDILOYRQOLKLOSTYIEGLKVTYRPSDKHVHTRFNOALTOGTGLSSTD 624
OY 703 ETWQDLVDVDFNGKRSOEALLAG-----VAHKTYADNLV-----NPMR 741
Db 625 PNLONIPIRLIEGKRIKQAFVPSKDWLIFADYSOIELRVLAHISKDENLIEAFTNDM- 683
OY 742 GYRORYSLEV-----GSSGLVSDANMAIARAGISGVYSGFDNAGSNRAHOMTGIOAGYIWS 799
Db 684 DIHKTAMDVYHAKDEVTSAMRQAKAVNFGIV-YGISDYGLSQNGITKEKAGAFI-- 740
OY 800 DNEHNVYRLRFPGAGDOSTIGYAHDSLPSIDKGYLTGGLVAGVAFVYNYEPEKDLRL 859
Db 741 DRY-----LESF-----OGVAYAMEDSVQAKOKGYVY-----TLMHRRYIPELTS 782
OY 860 AVFEDIGNAYDKGFTNDTKIGAGVGRWASPVQGVADVATGKEE 905
Db 783 RNFN-----IRSFARERTAMNTPIOGSAADILIKKAMIDMAKLEK 822

RESULT 32
DSG3 HUMAN
ID DSG3 HUMAN STANDARD: PRT: 999 AA.
AC P32926:
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Desmoglein 3 precursor (130 kDa pemphigus vulgaris antigen) (PVA).
GN DSG3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92069753; PubMed=1720352;
RA Amagai M., Klaus-Kovtun V., Stanley J.R.:
RT "Autoantibodies against a novel epithelial cadherin in pemphigus
RT vulgaris, a disease of cell adhesion."
RL Cell 67:869-877(1991).
CC -|- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
CC -|- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND
CC CARCINOMAS.
CC -|- DOMAIN: CALCINUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -|- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN
CC DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE
CC LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES
CC AGAINST DSG3.
CC -|- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M76482; AAA60230.1; -.
DR PIR: A41088; ITHUG3.
DR HSRP: P15116; INCUJ.
DR Genew: HGNC:3050; DSG3.
DR MIM: 169615; -.
DR InterPro: IPR002126; Cadherin.
DR Pfam: PF00028; cadherin; 4.
DR PRINTS: PR00205; CADHERIN.
DR SMART: SM00112; CA: 4.
DR PROSITE: PS00232; CADHERIN_1; 3.
DR PROSITE: PS0268; CADHERIN_2; 4.
DR Cell adhesion: Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.
KW SIGNAL 1 23
FT CHAIN 24 49
FT PROPEP 50 999
FT DOMAIN 50 615
FT DOMAIN 616 640
FT DOMAIN 641 999
FT DOMAIN 159 268
FT DOMAIN 269 383
FT DOMAIN 386 499
FT REPEAT 910 935
FT REPEAT 936 966
FT CARBOHYD 110 110
FT CARBOHYD 180 180
FT CARBOHYD 459 459
FT CARBOHYD 545 545
SO SEQUENCE 999 AA; 107503 MW; 60479DD46AC219A1 CRC64;

Query Match 2.7%; Score 128.5; DB 1; Length 999;
Best Local Similarity 19.9%; Pred. No. 15;
Matches 185; Conservative 129; Mismatches 377; Indels 239; Gaps 46;

OY 95 ISRIGESPRLGLDMSVLEETPLSLLEPLAQESTEMKINPDYIP-----EYGGQPRSE 150
Db 87 ISGVGIDPPPGI-FVVDKNTGDNITAIVDKEET-----PSFLITCALNAQG----- 134
OY 151 VVVPPTLEPERPGILKRLYARLFNDG--VNKVPRLAKAFYOSOSGETSATGSSHOKTER 208
Db 135 -----LDVEKP-LITVYKIIDINPNPVFSQIWMGEIENSASNSLWMLNATDADER 187
OY 209 -YANIKAALEDITQESAMDNLGSIPLRLQOTALVARAVGYDIDISTIRNSIGEVDYI 267
Db 188 NHLNSKIAFKIVSOEPA-----GTFWFLLSRNTGERTLTN 223
OY 268 DLGEVYVYIDYAVEVGSGADKAFIT-----VADEVPLLIGVFNHGKYEK-- 315
Db 224 SLDRQASSTLY-VSGADKQEGISTQCECNIKKYDVNDNPM-----FRDSQISATIE 277
OY 316 KNLENSAEGYFDGRWLDRSVDILPONTADVSLIVDTGTYQVFEDEVFEETIDPKTNQ 375
Db 278 ENILSSELL-----RQVYTDLDDEVYDNLAV-YFTSGNEGWMFEL---QTPRPTNE 326
OY 376 LTTDPDKLPVKREL-LBOLLVYNGEALNQLAVRLASDLATRFNVNVTIYPEREDQ 434
Db 327 -----GILKVVAKALDYEDQLOSVKLSIAVK---NKAEFQSVISRY-RVQSTPVYI--Q 373
OY 435 ION--DOVSFQSSSSRREPQAVDESTLEPIYEVVELLDGLMDISPIEFSASNLIOOKL 492
Db 374 VINNREGIAFRPAKST-----FTYQKGISSKLVYIL-----GYQAIDEDT 416
OY 493 NLVAKARHL-----YDMPDRVLAINHDDGVNRSILGRISDAVASAVARAI----- 538
```

```

Db 417 NKAASNNKYVGRDGGTGLMTDSTKAEIKYKNNMRDSTFVNTKTITAEVLAIDETGKT 476
QY 539 -----LPDESENEVIDLPERTALANRKTADYQOSKV-----PLYEYVASKPRD 584
Db 477 STGVYVVRPFPNCPNCPAIVLEKQAVCS-SSPSVVSARLTNNRTGP-YTFALDDQ-- 532
QY 585 GQIQLGMSDT--GTRLVTK-----FEHNLINRDGYOAGAEIRLSE-----DKK 626
Db 533 VKLPVAVSITLNTATSAIRAQEOIIPGVYHISLVLTDSQNNRCMPRSLFLEVQCQNR 592
QY 627 GVKLYATKPLSHP-----LNDOLKATFLGYQDEVFQHS 658
Db 593 GI-CGTSTPTTSPGTRRGPRHSRGLPRAIGLLLLGILLLLAPLLLTTCDCGAGSTGCV 651
QY 659 TNGF-----DLSTRLEHISRSIIIONGMNRTYSLRYRLDKLTKTO--APPETWODLPYDF 712
Db 652 TCGPIPVDPDSEGTI-HQ-----WG-IEGAHPEDKEITNICVPVT----- 690
QY 713 VNGKPSQDA-----LLAGVAHVHTYVADNLVNPVRGYRQRYSLLEVSGSLVSDANMAIA 765
Db 691 ANGADFESESEVCTNFTYARGTAVEGTSGEMETTKLGATESGGAAGFATGVSAASGFG 750
QY 766 RAGISGVYRCDNAYGSRARAHQMTGGIOAGY---IWSDNFNHVPYRLRFFAGGDSIRGY 822
Db 751 AATGVGICSSGQS--GTMRRHSTGCTNKDYADGAISMNFLDSYFSOKAFACAEEDDQOE 808
QY 823 AHDLSLPSDKY-LTGGQVLAVGTAEYNEFMKDLRLAVFGDIGNAVDKFTNTKIGA 881
Db 809 ANOCLLYDNMGADATSPGVSVCSCSFADDDLSFDLSIG-----PKFKLAETSL 861
QY 882 GVGVRMASPVGVQVRVDVATGVKEGKNPKTL 911
Db 862 GVDGE-GKEVOPRPSKDSGYGIESCGHPLEV 890

RESULT 33
RBP1_PLAVB STANDARD; PRT; 2869 AA.
ID RBP1_PLAVB STANDARD; PRT; 2869 AA.
AC 000798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN RBP1.
OS Plasmodium vivax (strain Belém).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT mercoites";
RL Cell 69;1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M88097; AAA29743.1; -
CC Malaria; Receptor; Signal; Transmembrane.
CC FT CHAIN 1 17
CC FT SIGNAL 17
CC FT CHAIN 18 2869
CC FT DOMAIN 18 2807 EXTRACELLULAR.

```

```

FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EB0FF CRC64;

Query Match 2.7%; Score 128.5; DB 1; Length 2869;
Best Local Similarity 18.9%; Pred. No. 75;
Matches 149; Conservative 127; Mismatches 302; Indels 209; Gaps 36;

QY 15 VALAIVLPLM-TSQL-----AQQNPANILNHVPHADAIN-----QAKAGNPVLLTP 63
Db 2180 ISLISYTANMKTSKRLIMINKERENTKCVDIKDNSSSTDGVETLKGFGYSKLTFFSSA 2239
QY 64 EQIQRLNAGIN-AKPOSALDVNFPDQSPISNIGESPPGLGDMVIE--ETTPSL 120
Db 2240 SETVONADPTSYNFAKKHEKESLNAIR-DIKKELYLFHQVS-----DISTVEGVONMLVL 2293
QY 121 BELFAQESTEMGINPDYIPEYQGEOPNSEVVPPTLEPEKPKLIRYLARLFNDGVNKV 180
Db 2294 YDKLINEERKREMDLYRNISETRKQMEHSTDFKRMIELHKG-----MNETNKS 2343
QY 181 PLKAKFYQSSQSGETSAIGSHQKTEPYANIKALLEDITQESAMDLSGSLPRLQOTALV 240
Db 2344 LLEKEKKLKV-----NDHMSMEAEIMKNGLK-YTPESYONINNI----- 2383
QY 241 AARAVGYDIDLSITRNSIGEVDYIHDGEPYIYDRAVEVRGSGADKATTVTADEV 300
Db 2384 -----YSVEAEVKTLEEDRDYGD--NYQIVE-----EHKKQPSILIDRTN 2423
QY 301 LLIGDFVHNGRYETKKNLE-NASAEHGYFDGRWLDLRSDVLLPNTADVSL-----YD 354
Db 2424 ALMDOI-BIFKKENNYNLMENYETIRHVND--YIEKTKNLKYQAKTEYEQLEENIKQND 2480
QY 355 TGTQYRDEEV-----FTIDPKTNQLTDT--PDKL-----PVKRELLQOLLTVNM 398
Db 2481 DMLQNIFFLKVSIIEFENVRKKKESIINDLYEQERLLKIGHEIRKKNVETL----- 2535
QY 399 GEAVNL-QAVVALSNDLATRYFNVNVEIYFPERE--QIQNDQVSFQSSSKRPEAOV 455
Db 2536 -SSTEIDQKMEMSKNLEKRSKMNNTYSIYLEKEANEIRNDAQIKDDDT-----I 2587
QY 456 DESTLEPIETVELTGLIDMIS-----PIEFSASNLIOOKLNV-----AAKARHLYDMPD 507
Db 2588 LNSVLEALQKRGMDALFQSMSADRNPNKYSAKYNNNEANEIIRQLEVLKRLREGOLVO 2647
QY 508 DRVLAINHDDVNSILGRISDAVASV-----ARAILPDESENEVIDLPERTALANRKT 562
Db 2648 D-----SESIISEMSSKKSALIEKERTARL--RTSENNRREERERARQEMSMN 2694
QY 563 ADVYQSKVPLVYVVASDKPRDQIGLGMSDGTGRVLYTKFEHNLINRDGYOAGAEIRLS 622
Db 2695 NDPQSEPT-----HSEGSIGEGKESDSDETGLT-----HDAGADEDST 2733
QY 623 EDKGVKLYATKPLSHPLNDLRAATLGYQGVFGSTNGFPLSTRLEHISRSIIQNG 682
Db 2734 SSAG-----AHELEEBETTAPEMETIRMDNTLLGID-TTRSDPDMHTENTQCG- 2782
QY 683 WNRITYSLRYRLDKLTKQAPPETWODL-----PVDFVNGKPS-----QEALL 723
Db 2783 -----TYQDTSNSDEADILNKKFNNVYKAGAFVLLCTSAVI 2821
QY 724 AGVAVHK 730
Db 2822 GATIAHK 2828

RESULT 34
FRPA_NEIMB STANDARD; PRT; 1302 AA.
ID FRPA_NEIMB STANDARD; PRT; 1302 AA.
AC Q9K0K9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```


16-OCT-2001 (Rel. 40, Last annotation update)
DE Iron-regulated protein frpA.
GN FRPA OR NMB0585.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.R.,
Hait D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
Rae G.M., Ullrich T.R., Khouli H., Qin H., Vamathevan J.,
Gill J., Scalapato V., Maignani V., Pizza M., Grandi G., Sun L.,
Smith H.O., Fraser C.M., Moxon E.R., Rappaport J.C.,
RA "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58".
RT Science 287:1809-1815(2000).
RL
CC -1- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL
DISEASE.
CC -1- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED (BY
SIMILARITY).
CC -1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC
CC EMBL: AE002414; AAF41013.1; -
DR TIGR; NMB0585; -
DR InterPro: IPR001343; HemIysn_Ca_bind.
DR Pfam: PF00353; hemolysinCbind; 12.
DR PRINTS: PR00313; CARNDRGPR.
DR PROSITE: PS00330; HEMOLYSIN_CALCIIUM; 9.
KW Toxin; Calcium; Outer membrane; Repeat; Complete proteome.
FT DOMAIN 752 975 13 X REPEATS, GLY-RICH.
FT REPEAT 761 766 2.
FT REPEAT 770 775 3.
FT REPEAT 789 894 4.
FT REPEAT 898 903 5.
FT REPEAT 907 912 6.
FT REPEAT 916 921 7.
FT REPEAT 925 930 8.
FT REPEAT 934 939 9.
FT REPEAT 943 948 10.
FT REPEAT 952 957 11.
FT REPEAT 961 966 12.
FT REPEAT 970 975 13.
SQ SEQUENCE 1302 AA; 141397 MW; 21D058C56C98BDE8 CRC64;
Query Match 2.7%; Score 126.5; DB 1; Length 1302;
Best Local Similarity 19.6%; Pred. No. 29;
Matches 184; Conservative 92; Mismatches 306; Indels 355; Gaps 41;
QY 163 GLIKRLVRLFN---DGNKKVRLAKFYOSSGSETSAIGSSHOKEPYA-----NIK 213
DB 223 GINDLYKSVKREWTGIFEIVNNNIKQFRDLFPNPEGWIDGHCQFAPWVKEFKRNKG 282
QY 214 ALEDITGESAMDLDGSIPLRLQGTALVAAVGY---YDIDLSIIRNSIGDV---YI 265
DB 283 YHYVD---PLADLDGD-----GIEIVAAKGFGSGLFDHTNNGIRIATGVAWSADGCLL 332
QY 266 IHDLGEPVYIDYRAVEVKGADKRAFTVADEVPLLIGDVFHNGKYETKKNLIENASAE 325

333 VRDLNGNGIID-----NGAE-----LFGD-----NTK--LADGSFAK 362
QY 326 HGCFDGRWLDRESDVLLPONTADVSL-----IYDTGQYRDEYVFFIID-- 370
DB 363 HGYYALAEIDSGNDIT--NAADAAFOSLRWODLNQDQISQANELRTLEELIGOSLDA 420
QY 371 -----PKTNOLTTD-----PDKLPVKRELLERLLT 395
DB 421 YKDVKNKLNGLNGNTLAQGSYTKTDGTAKMGDLLAADNLHSREKVKVELLEAOAKANL 480
QY 396 VMKGAEYVNIQAVRAISNDLIATRYFMVNTETVFPEREOIQNDQVSEFOSSSSRTPEAY 455
DB 481 AGIGRLRDJREAAALSGDLA-----NML-----KAYSAETKEAQL 516
QY 456 DSTLEPVETVELPDGILMDISPIEFA-----SNLIQKLNLVAA---K 498
DB 517 --ALDNLHKKAEIDSNMGKSPMLSTDWTQANEGIALTPSQVAQLKKNALVSLSDK 574
QY 499 ARHLYDMPDRLVAINHDDGVNRSILGRISDAVSAVARAILPDESENEYIDLPERTALAN 558
DB 575 AKAAIDAAARDRIAVLDAYTG-----QDSNTLYMSEEDALNI 611
QY 559 RKTPrADVYOSKVPPLV---VPVASDKPRDGOIGLGSGDT----- 595
DB 612 VVYVNDTDYHLAKNIYQNLLEFQTRLOPYLNQISFKMENDFTILDSGLVOAFENHVAETNP 671
QY 596 -----GTRLYVT-----KFEHN-----LINRDGY 613
DB 672 QKAFVDLAEMLAYGELRSWEGRLMTDVEEAKKAGKEDYOKVLGOETVALLAATSGT 731
QY 614 QAGAEIRL--SEDKKGVKIYATKP---LSHPLNDQLRATLGYQDEVFGHS-- 658
DB 732 QADDLIQNFGFGHNKKNVSYGNDGNDTLIGAGNDYLBEGSGSDTYVFGEGQDTVYNY 791
QY 659 -----TNGF--DLSTPTE--HEISR-----SIIONGMNRTSLRY 691
DB 792 DYATGKKDILIRFTDGTITADMLFTTREGNHLIKAKGSGQVYQSTFQNDG-----SGAY 846
QY 692 RLDKLKTQAPRETWODLPVDFVNGKPSQBALLAGVAHRTVADNLVNPGRYQRQRYSLV 751
DB 847 RIDE-----IHFDMGKVLDA-----TYKELVQGSTIDGSDRLYAYOS 883
QY 752 GS--SGLVSDAMKATRAAGISGYISFGDNAYGNSRAHQHTGGIOACTYISDNENHPRYL 809
DB 884 GMTLNGGLD-----DYLXGADGDDLNGDAGNDSTVSGN----- 918
QY 810 REFAGGDQSIIRGYAHDLSLPSIDKGYLTGGQVLAAGTAENYEFMKDLRLAVFGDIGNAY 869
DB 919 -----GNDTLDGGEQNDALGYNGNDALNGE-----GNDHLNGBDND---TLIGAGNDY 967
QY 870 -DKGFTNDTKT-GAGYGVMAWSPVGOVRYDVATGYKE 904
DB 968 LEGGSGSDTYVFGKFRG-----QDAVYNYDYATGRKD 999
RESULT 35
BCC2_ACETY STANDARD; PRT; 1326 AA.
AC 082861;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose synthase operon protein C precursor.
GN BCSC.
OS Acetobacter xylius.
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_TaxID=28448;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BPR 2001;
RX MEDLINE=98296257; PubMed=9630539;

RA Nakai T., Moriya A., Tonouchi N., Tsuchida T., Yoshinaga F.,
 RA Horinouchi S., Sone Y., Mori H., Sakai F., Hayashi T.:
 RT "Control of expression by the cellulose synthase (bcsA) promoter
 RL region from *Acetobacter xylinum* BPR 2001.";
 CC Gene 213:93-100(1998).
 CC -1- FUNCTION: Required for maximal bacterial cellulose synthesis. It
 CC may be involved in the formation of a membrane complex for
 CC extrusion of the cellulose product (by similarity).
 CC -1- PATHWAY: Bacterial cellulose biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ACSC/BCSC FAMILY.
 CC -1- SIMILARITY: CONTAINS 6 TPR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AB010645; BAA31465.1; -
 CC InterPro: IPR003921; CellSynth_C.
 CC InterPro: IPR001440; TPR.
 CC Pfam: PF00515; TPR.3.
 CC DR PRINTS: PR01441; CELLSTMTASEC.
 CC Cellulose biosynthesis: Repeat; TPR repeat; Membrane:
 CC Outer membrane: Signal.
 CC KW SIGNAL 1 30 POTENTIAL.
 CC FT CHAIN 31 1326 CELLULOSE SYNTHASE OPERON PROTEIN C.
 CC FT REPEAT 49 82 TPR 1.
 CC FT REPEAT 291 324 TPR 2.
 CC FT REPEAT 325 358 TPR 3.
 CC FT REPEAT 405 438 TPR 4.
 CC FT REPEAT 557 590 TPR 5.
 CC FT REPEAT 701 734 TPR 6.
 CC SEQUENCE 1326 AA: 142103 MW: FCAB2250C4E6C4A9 CRC64:
 SQ
 Query Match 2.7%; Score 126.5; DB 1; Length 1326;
 Best local similarity 19.0%; Pred. No. 30;
 Matches 199; Conservative 134; Mismatches 344; Indels 373; Gaps 52;

QY 506 -----PDDRYLAT-----NHDDGVNRSILGRISAVSAVARAILPDSENEVIDLPERT 554
 DB 536 LANPVTAODRQAGILYTGSGNDAMTROLAGLSPADYSPAIRSIAEMEIRQ-DLASRL 594
 QY 555 ALANKRTP-----ADYOSK-----KVPYLYVVASDKPD----- 584
 DB 595 SMVSNPVLIREALSQDPPTGARGVAVADLFRQGDVHARMAIRIASTRITDLSPPDRL 654
 QY 585 -----GOIGLWGSDDTGR-----VTRFENLJIND 611
 DB 655 SYATEYKISNPVAAARILALPGDGTSGAGNALLPEQMOTLQOLRMGISVAQSDLLNOR 714
 QY 612 GYQACA-----ELRLSEDKKGVKLYAT-----KPSHPPLNDLRLATL 648
 DB 715 GDQQAQYHILAPALQADEATSPKILARLYNGHGRPKKALEIDLAVLRHNPQDLARQA 774
 QY 649 GYQOEVEFHSINGEDLSRTRLEHISRSITIONGN-----NRTYS 688
 DB 775 AVQAAY-----NSNHSNLSATRLAMDGVQESPMDFARAWLMAVADQADHGQRTIEDLRRAVD 831
 QY 689 LRYRLDKL-----KTQAPPET-----MODLPVDVYNGKPSQEA--- 721
 DB 832 L-RLOQVEGTRAAAGPYGAHEALAPSTNPFOSRGYGHQVELGAPYTGGSYSAEASP 889
 QY 722 -----LLAGVA--VHKTVADNL-----VNPMDRGYRORYSLEVGSSGLVSDANMAIA-----R 766
 DB 890 DTSQMLSSINGOIR-TIRENLAPSIDGLGFRSR-SEHG-MGRULTEANIPYIRLPLQ 946
 QY 767 AGISGVSYFGDNAYGNSRAHOMTGGIOAGYIWSDFN-----HVPYRLRFAGGQDSIRG 821
 DB 947 AGASA-----LTFESITPTMWSGOLNTGSYVDV-RGTFEMATQANOC 989
 QY 822 YAHOSLSPISDKGYLTGS--QVLAVGTAEYNEEPKDLRLAVFGDIGAAYDKGFINDRI 879
 DB 990 AGHSSCGGLD--FLSANHTORIAAGAE-----AGFAPDVQF 1024
 QY 880 GAGGVNRW-----ASPVGQVRVDVATGVK 903
 DB 1025 GNS-----WVRADVCAASPIGPFITVNLGVE 1050
 RESULT 36
 FRPC_NEIMB
 ID FRPC_NEIMB STANDARD; PRT; 1829 AA.
 AC 09JVS:
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Iron-regulated protein frpc.
 GN FRPC OR NM01415
 OS *Neisseria meningitidis* (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; *Neisseria*.
 OX NCBI_Taxid=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / serogroup B:
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eissen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,
 RA Cotton M.D., Uutterback T.R., Knouri H., Qin H., Yamathayan J.,
 RA Gill J., Scarlato V., Maignani V., Piza M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappunli R., Venter J.C.:
 RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain
 RT MC58.";
 RL Science 287:1809-1815(2000).
 CC -1- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL
 CC DISEASE.
 CC -1- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED (BY
 CC SIMILARITY).

-1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
CC ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE002490; AAF41776.1; -
DR TIGR: NM81415; -
DR InterPro: IPR001343; Hemlysn_Ca_bind.
DR Pfam: PF00353; hemolysincabind; 22.
DR PRINTS: PR00313; CAMNDNGRPT.
DR PROSITE: PS00330; HEMOLYSIN_CALCIIUM; 17.
KW Toxin; Calcium; Outer membrane; Repeat; Complete proteome.
FT DOMAIN 879 1702 43 X REPEATS, GLY-RICH.
FT REPEAT 888 884 1.
FT REPEAT 888 893 2.
FT REPEAT 897 902 3.
FT REPEAT 1016 1021 4.
FT REPEAT 1025 1030 5.
FT REPEAT 1034 1039 6.
FT REPEAT 1043 1048 7.
FT REPEAT 1052 1057 8.
FT REPEAT 1061 1066 9.
FT REPEAT 1070 1075 10.
FT REPEAT 1079 1084 11.
FT REPEAT 1088 1093 12.
FT REPEAT 1097 1102 13.
FT REPEAT 1216 1221 14.
FT REPEAT 1225 1230 15.
FT REPEAT 1234 1239 16.
FT REPEAT 1243 1248 17.
FT REPEAT 1252 1257 18.
FT REPEAT 1261 1266 19.
FT REPEAT 1270 1275 20.
FT REPEAT 1279 1284 21.
FT REPEAT 1288 1293 22.
FT REPEAT 1297 1302 23.
FT REPEAT 1416 1421 24.
FT REPEAT 1425 1430 25.
FT REPEAT 1434 1439 26.
FT REPEAT 1443 1448 27.
FT REPEAT 1452 1457 28.
FT REPEAT 1461 1466 29.
FT REPEAT 1470 1475 30.
FT REPEAT 1479 1484 31.
FT REPEAT 1488 1493 32.
FT REPEAT 1497 1502 33.
FT REPEAT 1616 1621 34.
FT REPEAT 1625 1630 35.
FT REPEAT 1634 1639 36.
FT REPEAT 1643 1648 37.
FT REPEAT 1652 1657 38.
FT REPEAT 1661 1666 39.
FT REPEAT 1670 1675 40.
FT REPEAT 1679 1684 41.
FT REPEAT 1688 1693 42.
FT REPEAT 1697 1702 43.
SO SEQUENCE 1829 AA; 197445 MW; 8F63506E1FED9840 CRC64;

Query Match 2.7%; Score 126.5; DB 1; Length 1829;
Best Local Similarity 20.2%; Pred. No. 48;
Matches 186; Conservative 86; Mismatches 306; Indels 345; Gaps 40;

OY 170 ARLENDGVNKKVRLAKAKFYOSSGSETSAIGSSHOKTEPYA-----NIKALEDITQES 223
DB 361 AREFFKGLPSFKDLAEKFRDLFPNPE-GWIDDGHOCLAPWVETKRNKGKHYVD---PL 416

OY 224 AMDLNGSIPLRLQOTALVARAVGYDIDLSIRNSIGEY----VIHDLGEPYIIDRA 279
DB 417 ALDLGDD---GIEFAVATKGFAGSLDFDHTNNGIRATGWASADGLIVRDLNGNGIID--- 470
OY 280 VEVRREGADKRAFTYVADEVPLLLIGDVFHHGKVEFKKNLIENASAEHGFDGRWLDRSYD 339
DB 471 -----NGAE-----LFED-----NTR--LADGSFAKHGTAALAEILDSKD 503
OY 340 VILPNTAVSL-----IYDTGQYRDEVVEFTID----- 370
DB 504 NII--NAADAAQTILRVWQDNLQDQISQANELRTLELIGSLDLAYKDVKNKNGNTL 561
OY 371 -----PKTNQLTTD-----PKLYPKRLEQLLTVNNGEAVNLQAVRA 409
DB 562 AAGSVTKTKDTGTAKMGDLLLADNLHSRKKKVELTAAQAAANLAGLRDLREANA 621
OY 410 LSNDLIATRFYFMVMTVEIYFPEREQIQNDQVSFEQSSSRTPAQVDESLPVIETVEL 469
DB 622 LSGDLA-----NML-----KASAEETEAQL--ALDLNLHKMAE 655
OY 470 TDGIILMDISPIEESA-----SNLIQDKNLVAA--KARHLYMPDDRYIA 512
DB 656 TDSNMGKKSPPMLSTDWQTANEGIALTPSQVAQLKKNALVSLSDAKAIAIDARDRIAY 715
OY 513 INHDDGVNRSILGRISDANSAVARALPDSESENEVIDLPERTALANRKTPADYQSKYVP 572
DB 716 LDAYTG-----ODSNLYYMSSEDLNIVKYNTDITDHLAKN 752
OY 573 LY---VFVADKPRDQIGLQGWGSDT----- 595
DB 753 IYQNLLEFQTRLOPYLNQISFKMENDFTLDFSGLYQAFNHVKEETNPQKAFVDLAEMLAG 812
OY 596 -----GTRLYT-----KEEHN-----LINRDGQAQAERL--SEBK 625
DB 813 ELRSWYEGRRMLMTDYVEEAKKAKGKEDYQVLQETVALLARTSGTQADDILQNVGFGHN 872
OY 626 KGVNKLVIATKP-----LSHPLNDOLRATLGYOEFYFHS-----TN 660
DB 873 KVNSLYGNDSNDTLIGAGNDYLEGSGSDTYVFGSGFDQYVYNDYATGKRDIIRFTD 932
OY 661 GF--DLSTRPLE--HEISR-----SIIONGWNRTYSLARYLKLKTQAPETW 705
DB 933 GITADMLFTTBREGNHLILYKADSGGOVTVQSYFQNDG-----SGAYRIDE----- 977
OY 706 QDLPVDFVNGKSOEALLAGVAVHKTVAADNLVNPMMGYRQRTSLEVS--SGLYSDANMA 763
DB 978 ---IHFDMGKVLDTVA-----TVKELYQOSTDGSRLYAYQSGNTLNGSLGD----- 1020
OY 764 IARAGISGVYSFGDNAYGSRNAHOMTGGIOAGYIWSDNFNHVPYRLRFPAGDOSIRGYA 823
DB 1021 -----DLYXGADGDLLNGDAGNDSTISGN-----GNDTLGGSG 1055
OY 824 HDLSLPSIDKGYITGGQVLAIVTAENYFEMKDLRLAVGDIQNAV--DKGFTNDTRI--GA 881
DB 1056 NDALYGYNGNDALNGGE---GNDHLNGEGDND---TLIGAGNDYLEGSGSDTYVFGK 1108
OY 882 GYGVRWASPVGQVRVDYATGVKE 904
DB 1109 GFGQDVTY-----YNDYATGRKD 1126

RESULT 37
RPL_HUMAN
ID RPL_HUMAN STANDARD: PRT: 2156 AA.
AC P56715;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Oxygen-regulated protein 1 (Retinitis pigmentosa Rpl protein)
DE (Retinitis pigmentosa 1 protein).
GN RPL OR ORPL.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS HIS-872; TYR-985; THR-1670; PRO-1691
RP AND TYR-2033.
RC TISSUE=Retina;
RX MEDLINE=99318096; PubMed=10391212;
RA Sullivan L.S., Heckenlively J.R., Bowne S.J., Zuo J., Hide W.A.,
RA Gal A., Denton M., Inglehearn C.F., Blanton S.H., Daiger S.P.;
RT "Mutations in a novel retina-specific gene cause autosomal dominant
RT retinitis pigmentosa";
RL Nat. Genet. 22:255-259(1999).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Retina;
RX MEDLINE=99318095; PubMed=10391211;
RA Pierce E.A., Quinn T., Meenan T., McGee T.L., Berson E.L., Dryja T.P.;
RT "Mutations in a gene encoding a new oxygen-regulated photoreceptor
RT protein cause dominant retinitis pigmentosa";
RL Nat. Genet. 22:248-254(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99330563; PubMed=10401003;
RA Guillonneau X., Pirliev N.I., Danciger M., Kozak C.A., Cideciyan A.V.,
RA Jacobson S.G., Farber D.B.;
RT "A nonsense mutation in a novel gene is associated with retinitis
RT pigmentosa in a family linked to the RPI locus";
RL Hum. Mol. Genet. 8:1541-1546(1999).
CC -1- FUNCTION: COULD HAVE A ROLE IN THE DIFFERENTIATION OF
CC PHOTORECEPTOR CELLS.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN RETINA. NOT EXPRESSED IN HEART,
CC BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, SPLEEN AND
CC PANCREAS.
CC -1- DISEASE: DEFECTS IN RPI CAUSE RETINITIS PIGMENTOSA FORM 1 (RPI); A
CC DISEASE CHARACTERIZED BY CONSTRICTION OF THE VISUAL FIELDS, NIGHT
CC BLINDNESS, AND FUNDUS CHANGES. THE DISEASE SEEMS TO BE ASSOCIATED
CC WITH TRUNCATED (STOP OR FRAMESHIFT MUTATIONS) FORMS OF THE
CC PROTEIN.
CC -1- SIMILARITY: CONTAINS 2 DOUBLECORTIN DOMAINS.
CC -1- DATABASE: NAME=Retnet;
CC NAME=Retinal Information network;
CC WWW="http://www.sph.uth.tmc.edu/retnet/".
CC -1- DATABASE: NAME=Mutations of the RPI gene;
CC NOTE=Retina International's Scientific Newsletter;
CC WWW="http://www.retina-international.com/sci-news/rpimut.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF143226; AAD44197.1; -;
DR EMBL: AF143224; AAD44197.1; JOINED.
DR EMBL: AF143225; AAD44197.1; JOINED.
DR EMBL: AF143223; AAD44198.1; -;
DR EMBL: AF141021; AAD42072.1; -;
DR EMBL: AF152242; AAD46774.1; -;
DR EMBL: AF152240; AAD46774.1; JOINED.
DR EMBL: AF152241; AAD46774.1; JOINED.
DR EMBL: AF146592; AAD46769.1; -;
DR Gene: HGNC:10263; RPI.
DR MIM: 603937; -;
DR MIM: 180100; -;
DR InterPro: IPR003533; DCX.
DR Pfam: PF03607; DCX; 2.
DR SMART: SM00537; DCX; 2.
DR PROSITE: PS03039; DC; 2.
KW Vision; Retinitis pigmentosa; polymorphism; Repeat.
FT DOMAIN 36 118 DOUBLECORTIN 1.

FT DOMAIN 154 233 DOUBLECORTIN 2.
FT DOMAIN 268 273 POLY-SER.
FT DOMAIN 671 675 POLY-LYS.
FT DOMAIN 1687 1691 POLY-SER.
FT DOMAIN 872 872 R -> H.
FT VARIANT 985 985 /FTid=VAR_007810.
FT VARIANT 1670 1670 N -> Y.
FT VARIANT 1670 1670 /FTid=VAR_007811.
FT VARIANT 1691 1691 A -> T.
FT VARIANT 1691 1691 /FTid=VAR_007812.
FT VARIANT 2033 2033 S -> P.
FT VARIANT 2033 2033 /FTid=VAR_007813.
FT VARIANT 2033 2033 C -> Y.
FT VARIANT 2033 2033 /FTid=VAR_007814.
SQ SEQUENCE 2156 AA; 240659 MW; 55ADEBC43D6A507 CRC64;
Query Match 2.7%; Score 126.5; DB 1; Length 2156;
Best Local Similarity 17.5%; Pred. No. 62;
Matches 181; Conservative 159; Mismatches 367; Indels 325; Gaps 42;
QY 22 PLMTSQAALQONNPNANITNHPADHTALNQAKGNPVLTPEDIQARLNAAGLNAPQOS 81
DB 126 PMLSSRAISA-----HSPHPVAV-----AAGMPRPPRS 155
QY 82 QALDVNFDQSPISRIQEQSPPIGLQMSYTEHTPLSELEFQESTEMGinp----- 135
DB 156 ----LVYFRNDPDKTR-----RAVLSSRRVYQSFAPLQHLTEVMQRPVVKLYA 200
QY 136 --NDYIPEYQGEQPNSEVVPPTLEPEKPGILKRLYLAFNDGVNK-----VPLKAKF 187
DB 201 TDGRVPSLAVALISSGAVNAAGREPKPG-----NYDIQKYLPLARLGISORV 250
QY 188 YQSOSGETSAIGSHQKTEPYANIKALEDIQESADLNGSIPLRLQALVA--ARAV 245
DB 251 YPKGNASESKISTHSSSSRSQIYSSVSEKTHNNOCYLDYSEVPEKYLALERNDSQNL 310
QY 246 GYV----DIDSIIRNSIGEVDYIIHDLGEVYIDYRAVEFRGAGDADKAF--TYVA-- 296
DB 311 PIYSEDDIEKSTIFPNQDGTWT-----EMKRFRIKEETIKMTTTSKT 356
QY 297 ----DEVPLIGDVFHNGKYEYTKK-----NLINASAEHGYFDGRMLDRS 337
DB 357 GPSNNDK-----SEMSPGRTESSSGKLACFSADVSMEMSSNOEGS-----LAEE 407
QY 338 VDYLPPNTADVSLIYDTGYQRFDEVVFTIDRKTQLTDPDKLPVKRELBOLLTVN 397
DB 408 INIQMTQVA-----ETCSASWENA---TVPDIIQGDQAKHFRYRPPRGLRVR 458
QY 398 MGEAYNQAVRALSNDLIATRYFMVTELYFPEREIQNDQV-----SFEQSSSRTE 451
DB 459 QKRSV-IGSVTLVSETEVOEK--MIQFYSSEEREGENKSEYHMFTHSCSKMSVSNK 514
QY 452 PA-----QVDESTLEPTE-----TYELTGILMDIS-----PIEFASN 486
DB 515 PVLVOINNDQMESSLEKKNLSKSAISAGVIEITISQKMLKMSHNNGLDSTISNNS 574
QY 487 LIDDKL-----NLVAAKARHLVDPDDRVLAINH-----DGVNRSI-----LGRI 527
DB 575 IVEEDVVDVLDNKTGIGIKFKFTYGTNDRFSPISADATHFSSNNGSTDKNISEAPASEA 634
QY 528 SDVASAVARAILPDESENEVIDLPKRALANRKTTPADYOSKVPVLYVYASAKPRDQI 587
DB 635 SSTVTARIDRLNEFAQCGLTKLPKN---EKKILSSVASKKRRKSROAINSRYODGQL 690
QY 588 GLGWSGDTGRLVTK--FEHNILNRD-----GYQAGAEPLRLSE----- 623
DB 691 ATKGIILKNMERINTKGRITKEMIVQSDSPILKGCILCEEDLQASDVYIESNTFCSSNLN 750
QY 624 -----DKGKVLVATKPLSHPLND-----QLRATLGYQOEVFGSHSTNGFD 663
DB 751 STISKNFHNRKRLTQNSKVGGLLTRKRSRLNKISLGAAPKKEIQGRQKVFPHN----- 805
QY 664 LSTRLEHLSIISIIQNGMNRITSLRRLDKLKTQAPPETWODLPVDVYVNGKPSQDALL 723

```

Db 806 -----ESKCYKSTFEENKSLFHFVFIIEQ--KPKDFYAPQSOAEVAGSYLRG-----M 850
Oy 724 AGVAVHKVAVNLVPMGYRORYSLEVSSGLVS-----DANMAIAR 766
Db 851 AKKSILVSKVTOSSHIT-LKSOKKRRKDKYKASAILSKOHATTRANSLSLKKPFPEALIH 909
Oy 767 AGI-----SGVYSFGDNAYGNSRAHOMTGGIOAGYI 797
Db 910 HSIQNYIOWLQININPYPTLKRPIKSAFYCRNETSVYVNCNSNFSGNDPHTNSGKI-SNFFV 968
Oy 798 WSDNHNHYPRLRFRFAGGDSIRGVAHDSLSPISDKGLTGGQVLAAGTAENYEFEMDL 857
Db 969 MESN-KHI-----TKIAGLTGDNLCKEGDKSF-----IANDTGE-----EDL 1004
Oy 858 RLAVFGDIGNAV 869
Db 1005 HETGVGSLNDAY 1016

RESULT 38
YLJ2-CAEEL STANDARD; PRT: 2198 AA.
ID YLJ2-CAEEL STANDARD; PRT: 2198 AA.
AC P34367;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 256.3 kDa protein C50C3.2 in chromosome III.
GN C50C3.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RA [1]
RA SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardiner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Lareille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shawkeen R.,
RA Sims M., Smailson N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RA elegans."
RA Nature 368:32-38(1994).
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: L14433; AAA27973.1;
CC WormPep: C50C3.2; CE01861.
CC InterPro: IPR002048; EF-hand.
CC InterPro: IPR002017; Spectrin.
CC Pfam: PF00036; EFhand; 2.
CC Pfam: PF00435; Spectrin; 10.
CC ProDom: PD000012; EF-hand; 1.
CC SMART: SM00150; SPEC; 5.
CC PROSITE: PS00018; EF_HAND; 1.
KW Hypothetical protein. Calcium-binding; Repeat.
KW CA_BIND 2025 2036 EF-HAND (POTENTIAL).
FT SEQUENCE 2198 AA; 256260 MW; CDB2C1092C5EDC8 CRC64;

```

```

Query Match 2.7%; Score 126.5; DB 1; Length 2198;
Best local similarity 18.7%; Pred. No. 64;
Matches 124; Conservative 101; Mismatches 248; Indels 189; Gaps 27;

Oy 81 SGALDVVNFDDSPISR--IGROSPPLGLDMVIEETPLSLLELFAESTMGINPDY 138
Db 1232 NDSMSQIILDEQFDITRQWKRTALEADVYSKTRPDIERM-QKRHDELAJDINDY 1290
Oy 139 IPEYGEQPNSEVVPPTLEPEKPGILKRLVRLFNDGVNKPRLKARFYOSSQ-----S 193
Db 1291 HTIYEDFLNNGVY-----EEEKVQVIRHLWSSLLESSTTRQKSL-AQCQKSRFLFELD 1343
Oy 194 GFTSAIGSSHQKTEPYANIKALLEDITQESAMDNGSIPRL--KQTAIVARAAGYDI 250
Db 1344 GMSIWLIDAEFDVHSTTFEIGLYNSVDAEKARLLVSLGEQAGEKDDILIRIQCESVDY 1403
Oy 251 DLSIRNSIGEVDYTIHHDGEPVYIDYRAVEYRGSGADKAFITVADEVPLL--IGDVFH 308
Db 1404 IYERLRNGISQIOMLIKANNEDLV-WKSMQ-----KYVTAIDDELCEWKEINVIYS 1454
Oy 309 HGKY--ETRKNLINENASAEHGYFDGRWLDKRSVDVILPDNTADVSLIYDTGYRDEYVF 366
Db 1455 STNVGNDVSSNDVLRKRKHQRLQLEFORQOKVAKVYLLTTELVSRRPSLEYKFEV-- 1511
Oy 367 FTIDPKTQNTDPPKLPYKRELLFQLLTVNNGEAVNIQAVNALNDILA--TRYENMVT 425
Db 1512 --IDKRVAVLT-----ELLSS-----NRQIAIRIKRLEKWTETVYVME 1549
Oy 426 EIVPFEREQIQNDQVSEFQSSSRTEPAQVDESTLEPVLETVDLGLMDISPIEFSAS 485
Db 1550 ---IREKEGLL-DQLLDLKTGTPEIVLADVERKL--QVLETV----- 1585
Oy 486 NLIQKLNIVAAKARHLIDMPDDRYLAINHDGVNRSILGRISDAVSAVARAILDESEN 545
Db 1586 ---GDHMDTLITKAE---QMSDDEVY-----RTKVAVTA-----DQLRK 1619
Oy 546 EVIDPERALANRKPADVYQSKKVPYLVFVAPSDKPRDQIGLMSGSTGRLYTKFEB 605
Db 1620 KVLKNNELKKHHQIKESIETK-----FVST-----CDTGIOCI----- 1655
Oy 606 NLINNDGYQAGAEELRLSEDKKGVKLKATKPLSLPINDQIRATL---GYQGVF----- 655
Db 1656 -----REQEEKIGNLVHOKPASPENFENTAYHSTFVEYTKDFFELKVY 1701
Oy 656 -----GHSTNG-----FLLSTRTEHEHSRTIQNG 681
Db 1702 CRKIQNSTDAEKRLSLVSERLNALKKQDLDLAEKIAVDKFKLVQNIODEYSRTACEIG 1761
Oy 682 GW 683
Db 1762 NW 1763

RESULT 39
PCGV_HUMAN STANDARD; PRT: 3396 AA.
ID PCGV_HUMAN STANDARD; PRT: 3396 AA.
AC P1611; P20754; Q9UNW5; Q13010; Q13189; Q15123;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Versican core protein precursor (large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glia)
DE hyaluronate-binding protein) (GHA).
GN CSB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RA [1]
RA SEQUENCE FROM N.A. (ISOFORM VO).
RA MEDLINE=95105188; PubMed=7528742;
RA Naso M.F., Zimmermann D.R., Iozzo R.V.;

```

RT "Characterization of the complete genomic structure of the human
RT versican gene and functional analysis of its promoter.";
RL J. Biol. Chem. 269:32999-33008(1994).
RN (2)
RN SEQUENCE FROM N.A. (ISOFORM V1).
RC TISSUE=placenta;
RX MEDLINE=90059882; PubMed=2583089;
RA Zimmermann D.R., Ruoslahti E.;
RT "Multiple domains of the large fibroblast proteoglycan, versican.";
RL EMOB J. 8:2975-2981(1989).
RN (3)
RN SEQUENCE FROM N.A. (ISOFORM V2).
RC TISSUE=glial tumor;
RX MEDLINE=95105187; PubMed=7806529;
RA Dours-Zimmermann M.T., Zimmermann D.R.;
RT "A novel glycosaminoglycan attachment domain identified in two
RL alternative splice variants of human versican.";
RL J. Biol. Chem. 269:32992-32998(1994).
RN (4)
RN SEQUENCE OF 2711-3396 FROM N.A.
RC TISSUE=Lung fibroblast;
RX MEDLINE=88007514; PubMed=2820964;
RA Krusius T., Gehlsen K.R., Ruoslahti E.;
RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
RL lectin-like and growth factor-like sequences.";
RN J. Biol. Chem. 262:13120-13125(1987).
RN (5)
RN SEQUENCE OF 251-347 FROM N.A.
RX MEDLINE=93122792; PubMed=1478664;
RA Tozzo R.V., Naso M.F., Camilizaro L.A., Wasmuth J.J.,
RA McPherson J.D.;
RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
RL human chromosome 5 (5q12-5q14).";
RN Genomics 14:845-851(1992).
RN (6)
RN SEQUENCE FROM N.A. (ISOFORM V3).
RC TISSUE=Brain;
RX MEDLINE=95181355; PubMed=7876137;
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RL without a chondroitin sulfate attachment in region in mouse and human
RN tissues.";
RL J. Biol. Chem. 270:3914-3918(1995).
RN (7)
RN SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
RC TISSUE=Aortic smooth muscle;
RX MEDLINE=99327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA Wright T.N.;
RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN (8)
RN PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=89174663; PubMed=2466833;
RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
RT "Isolation and partial characterization of a glial
RL hyaluronate-binding protein.";
RL J. Biol. Chem. 264:5981-5987(1989).
RN (9)
RN TISSUE SPECIFICITY OF ISOFORMS.
RX MEDLINE=96213482; PubMed=8627343;
RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
RT "Differential expression of versican isoforms in brain tumors.";
RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
RN (10)
RN FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC (11) SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC (12) ALTERNATIVE PRODUCTS: At least 5 isoforms; V0 (shown here), V1,
CC V2, V3 and Vint; are produced by alternative splicing.
CC (13) TISSUE SPECIFICITY: Cerebralwhite matter. V0 and V1 is expressed

```

CC      in normal brain, gliomas, medulloblastomas, schwannomas,
CC      neurofibromas, and meningiomas; v2 is restricted to normal brain
CC      and gliomas; v3 is found in all these tissues except
CC      medulloblastomas.
CC      -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC      -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC      -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC      -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC      -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC      -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC      -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U15936; AAA65018.1; -.
DR      EMBL; X15998; CA34128.1; -.
DR      EMBL; S52488; AAB24878.1; -.
DR      EMBL; U26555; AAA67565.1; -.
DR      EMBL; D32039; BAA06801.1; -.
DR      EMBL; J02814; AAA36437.1; -.
DR      EMBL; AF084545; AAD48545.1; -.
DR      PIR; S06014; S06014.
DR      PIR; A29348; A29348.
DR      PIR; A30358; A30358.
DR      HSSP; P01132; 1EGF.
DR      Genew; HGNC:2464; CSPG2.
DR      MIM; 118661; -.
DR      InterPro: IPR000152; Asx_hydroxyl.
DR      InterPro: IPR000561; EGF-like.
DR      InterPro: IPR00742; EGF-2.
DR      InterPro: IPR001881; EGF_Ca.
DR      InterPro: IPR001438; EGF_II.
DR      InterPro: IPR003599; Ig.
DR      InterPro: IPR003006; Ig_MHC.
DR      InterPro: IPR001304; Lectin_C.
DR      InterPro: IPR000538; Link.
DR      InterPro: IPR000436; Sushi_SCR_CCP.
DR      Pfam; PF00008; EGF; 2.
DR      Pfam; PF00047; Ig; 1.
DR      Pfam; PF00059; lectin_C; 1.
DR      Pfam; PF00084; sushi; 1.
DR      Pfam; PF00193; Xlink; 2.
DR      PRINTS; PR0010; EGFBL00D.
DR      ProDom; PD000918; Link; 2.
DR      SMART; SM00032; CCP; 1.
DR      SMART; SM00034; CLECT; 1.
DR      SMART; SM00179; EGF_CA; 1.
DR      SMART; SM00001; EGF-like; 1.
DR      SMART; SM00449; IG; 1.
DR      SMART; SM00445; Link; 2.
DR      PROSITE; PS00010; ASX_HYDROXYL; 1.
DR      PROSITE; PS00022; EGF_1; 2.
DR      PROSITE; PS01186; EGF_2; 1.
DR      PROSITE; PS01187; EGF_CA; 1.
DR      PROSITE; PS01241; Link; 2.
DR      PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR      PROSITE; PSS0041; C-TYPE_LECTIN_2; 1.
KW      Glycoprotein, Proteoglycan, Lectin, Extracellular matrix, Sushi,
KW      Signal, Repeat, EGF-like domain, Calcium, Immunoglobulin domain,
KW      Hyaluronate acid, Alternative splicing.
FT      SIGNAL 1 20
FT      CHAIN 1 3396
FT      DOMAIN 37 137
FT      DOMAIN 167 244
FT      DOMAIN 265 346
FT      DOMAIN 348 1335
FT      DOMAIN 348 1335

```

```

FT DOMAIN 1336 3089 GAG-BETA.
FT DOMAIN 3089 3125 EGF-Like 1.
FT DOMAIN 3125 3163 EGF-Like 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3163 3290 C-TYPE LECTIN.
FT DOMAIN 3290 3353 SUSHI.
FT DISULFID 44 130 BY SIMILARITY.
FT DISULFID 172 243 BY SIMILARITY.
FT DISULFID 196 217 BY SIMILARITY.
FT DISULFID 270 345 BY SIMILARITY.
FT DISULFID 294 315 BY SIMILARITY.
FT DISULFID 3093 3104 BY SIMILARITY.
FT DISULFID 3098 3113 BY SIMILARITY.
FT DISULFID 3115 3124 BY SIMILARITY.
FT DISULFID 3131 3142 BY SIMILARITY.
FT DISULFID 3136 3151 BY SIMILARITY.
FT DISULFID 3153 3162 BY SIMILARITY.
FT DISULFID 3169 3180 BY SIMILARITY.
FT DISULFID 3197 3289 BY SIMILARITY.
FT DISULFID 3265 3281 BY SIMILARITY.
FT DISULFID 3296 3339 BY SIMILARITY.
FT DISULFID 3325 3352 BY SIMILARITY.
FT CARBOHYD 57 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 809 809 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1332 1332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1338 1338 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1442 1442 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1468 1468 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1663 1663 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1898 1898 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2179 2179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2272 2272 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2280 2280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2360 2360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2385 2385 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 2.74; Score 126.5; DB 1; Length 3396;
Best Local Similarity 19.8%; Pred. NO. 1.2e+02;
Matches 153; Conservative 118; Mismatches 293; Indels 207; Gaps 39;

66 IOARNAGLNAKPO---SQALDVNFPDQSRISIGOSPGLGDMVIEET--PLSL 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
358 ILAEFASLSLKEPQWMDRTPIILVDELPI--IPREFPVGNIVSEQKATVQPOAI 415
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

121 EELFAOE-----STEMGINPDYIPYOGEOPNSEVVVPTLEPEKGLKRLYARLFN 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

416 TDSLATKLPPTGTSTKPKWMDMDVSPSASGPL----- 447
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

175 DGVNKPPLKAKFVQSSGSEISAGSHOKPEPYANI---KALEDITQ-----E 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

448 -GKLDISEKKEVLDST-----TGVSHYATDSWDGVYEDKQTESYQIQIEVGPLV 499
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

223 SAMDLNGSIPR---LRQATVAVARAVGYIDLSIRNSIGEVDYI--DLGEPYVI 275
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

500 TSMELKHPSPKEPVTEPLVTARMILESKTEKEMV--STVELVTTGHYGLTEDEDE 558
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

276 DYRAVEVGEQADKAFITVADVPULL-----GDVFHNGKTEYKKNLIENASHEGTF 329
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

559 D-RTLTV--GSDS--LFIPOQIEVITVSKTSEDTIH-----THLEDELSVAS-- 603
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

330 DGRWLDNRSDVILPNT--ADVSLIVDTGTVRFDEVV--FFTIDPKTQTLTDDKLPV 385
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

604 -----TYSPLIMPONNGSSMDMWERQTSGRITTEEPGLKYISTPFPQHTETIEELFPY 658
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

386 KRE--LEQLTV-----NGEAYNLQAVRALSNDLIATRYNNMVTETV--FPERQION 437
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

659 SGDKLIVGIGISTVYIPSLQTEMTTHRERTETLIPKRDYTVDEIOELETKSPFMKTEE 718
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

438 DVSPFEQSSSTPEPAOVDESFLPEVETVELTDLGILMDISPIESASNLIDKLVLVA 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

719 EVFSGMKLSTSLSEPIHYTES-----SVEMTK-----SDFPITLI---TKLSA 758
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 498 KARHLYMPDDRVL---AINHDDGVNNSIL--GRISDAVSAVARAILPDESENVIDLP 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 759 EEPVREMRMEDEFATPGTGTCKDENITTVLHAGTSLVEAAVSKWSW--DEDNNTSKPL 817
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 553 RTALANRKPADYV-----QSKKVPYVVASDK-----PRDQIGLGMSDPTGRVLT 601
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 818 TEPASASKLPALLTGYGMKNCKDKDIPSEFDGADDEFILPDSQOKOLEVTDIDIAHG 877
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 602 KF--EHNILINDGYQAGAEIRLSDDKGV-----KLVTAKPLSHPLNDQPATIGYQ 651
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 878 KFTIRFOPTSTGLAEKSTLSDSTTEEKVPIPTSTEGQVATMGSS-----ALGEV 928
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 652 QEV-----FGHSTN-----GFDLSTRTLE-----HEISRSIIONGWNRTYS 688
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 929 EDVLDKRPVSTVPOFAHTEVEGALFVSYSTOEFTTVYDSSHITPLSVIRKTDWG----- 984
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 689 LRVRLDKLTQAPPETWODLPVDFVNGKPSOBALAGVAVHRTVADNLVNP 739
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 985 -----VLVPSVPSF-----DEVLGERSQDILV-----IDOTRLEATVISP 1018
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 40
ID P200_MYCGE STANDARD; PRT: 1616 AA.
AC 049429; 049259; 049298; 049352; 049353;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein P200.
GN M3386.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kiehl J., Smit A., Smit A.,
RA Fritchman J.L., Weidman J.F., Smell K.V., Sandusky M., Fuhrmann J.L.,
RA Tomb J.-F., Dougherty B.A., Bolt K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RA "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 256-427; 432-543 AND 1083-1140 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bolt K.F., Hutchison C.A. III;
RA "A survey of the Mycoplasma genitalium genome by using random
RA sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -I- FUNCTION: COULD BE A ACCESSORY STRUCTURAL COMPONENT IN
CC CYTADHERENCE (BY SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
CC EMBL: U39720; AAC71613.1; -
CC EMBL: U02245; AAA03400.1; -
CC EMBL: U02245; AAA03401.1; ALT_FRAME.
CC EMBL: U02175; AAD12458.1; -
CC EMBL: U02126; AAD12402.1; -
CC TIGR: M3386; -
CC Cytadherence; Structural protein; Repeat; Complete proteome.
FT DOMAIN 1205 1389 2 X 32 AA REPEAT.

```

```
FT REPEAT 1205 1236 1-1.
FT REPEAT 1358 1389 1-2.
FT DOMAIN 891 1389 2 X 26 AA REPEAT.
FT REPEAT 1161 1186 2-1.
FT REPEAT 1310 1339 2-2.
FT CONFLICT 256 256 P -> S (IN REF. 2).
FT CONFLICT 304 304 S -> F (IN REF. 2).
SQ SEQUENCE 1616 AA: 185678 MW: 6AF76A13AC49E4FF CRC64;
```

Query Match

2.7%; Score 125.5; DB 1; Length 1616;

Best Local Similarity 19.5%; Pred. No. 46; Mismatches 180; Indels 183; Gaps 31;

Matches 111; Conservative 94; Mismatches 180; Indels 183; Gaps 31;

```
QY 76 NAKPQSGALDVNFDDSPISRGQSPPLGLDMSVIEETPLSLLELFAQESTMG--- 132
DB 297 NTKSLASSL-----ENLLLENSSEQP-----VIEVAPRRNEVIFRNPVTKLHEK 344
QY 133 -----INP-----NDYIP--EYQEQ--PNSVYVPPPLEP---KPL---IKRLYARL 172
DB 345 EKEEFLNPVKETETIPLIIVKEEVESEVEAPVEIEPEACEPETIPEVETVF--V 402
QY 173 FNDGV-----NKVPRLKAF--YQSQSGEISAIGSSHOKTEPYANIKALEDITQ 221
DB 403 YEDDLKGLDGNQOAGNVPEVEYFVEDLKGLDSTIKDDQOHDE---IAKHVEHLSQ 458
QY 222 ESAMDINGSIPRLQGTALVARAVGYIDLSIRNSIGEV-----DVIIDHIGEPV 273
DB 459 DYSKEIKDS-----AKA-----DLSNISDDIDSVWKEFGSFTDETOKSVEKS 501
QY 274 YIYRAVEVGEQADKAF--TTVADEVPLLIGVFHHGKYET--KKNLIENASAEHGVPD 330
DB 502 QVDEIILDANNDEINSLFRDEVVNNIDSOINETVSEQEFPYVSNEFOEFSEPVSD 561
QY 331 GRW-----LDRSV-----DVILPDNTADVSLIYDCTOYR-----FD 362
DB 562 EKIKETNSDESVTDTALFSEKLVNEVLLTNEYVDVNAFPSTETEVKVSSELPKSELVD 621
QY 363 EVVFTIDPKTNQ-----LFTDPDKL-----P-VKRELLEQLLTVN-MG 399
DB 622 EITFFINNDPQEGLEKVDLFLETEPKSLPDEKTTIVSESEPPFIQPDLSLELDVNDVD 681
QY 400 EAVNIQAVRA--LSNDLIATRYFNM-----VNTETIYPEREQIONQVSF 442
DB 682 KSLKETTSVELNHEEIGNEFINLDVSEKEVQOPTOLETDSEFVLPYQIIVED---SF 738
QY 443 EOSSSRTEPAQVDESTLEPVIEFVELTDGILMDISPIESASNL-----IQDK 491
DB 739 TESAEIPNEFSSSQKDTLEFISQTOE-----VETSESNAVPTVEQETKLFERHODE 787
QY 492 LNLVA-----AKARHLYD-MPDDR 509
DB 788 NNLFTPLPLDLFEIIESNALFDSKPEDK 815
```

Search completed: April 28, 2003, 16:22:31
Job time : 67 secs


```
Db      87 VKDGKPKLTLKVVYGEFRLROVNIQVLEGAASLESFRLPSGK -OLKPGAKLNGVYED 145
Qy      315 KKNLEMSAHEGYFDGMDLRSDVILIPDNFADVSLIYDGTQYRPEVEYFTIDPKTN 374
Db      146 AKRLIONASRGFFQGRSTORLSDIPRAGIADIDLVDYSQRTTFGKVSF----- 197
Qy      375 QLTDPDKLPVKRELLECLLTVMNGEAYNLQAVRALSNDLIATRYFNNVNTETIVEPEREQ 434
Db      198 ----DGDST -IEELLRRWVPFKAGQPYDSIELIENQNLQSGSF----- 238
Qy      435 IONQVSEEOSSSRTPEAQNDESTLEPIETVELTDGILMDISPIESASNLIDDKLNL 494
Db      239 -----EGVRDAAPTQQA----- 252
Qy      495 VAAKARHLDMPPDRVLA1NHDDGVNRSILGRISDAVASAVARAILPDESENEVIDLPERT 554
Db      253 -----DGAQA----- 259
Qy      555 ALANRKTADVYQSKKVPVLYFVASDKPRDQIGLGWGSDTGTRLVTKFEHNLINRDGYQ 614
Db      260 -----PVAVRLERAKPRMTMGVGLGFSTDVGARARFNTRHVNAEGHS 302
Qy      615 AGAELRLSEDKKGVKLYTKPLSHPLNDLRLATLGQOEVEFHSINGPDLSTRTL----- 670
Db      303 LGFESEISAPRONVGAMYEIPLDPLTDKRLRTSGTQFE-----DLVDTESKLTLTGGSE 356
Qy      671 HEISRSIIIONGMNRTYSLRYRLDKLTKQAPEPTWQDLPVDFVNGKPS--QOALLAGVAV 728
Db      357 WHSKRP-----DCMQRVVSLMMREYKL-----GDDSGLSFLMPGICY 396
Qy      729 KKTVDNLVNPFRGYRQKRSLEVGSSGLVSDANMALIARAGISGVYSGFNATVGSRAHOM 788
Db      397 SLETDNKNVDPHGRYGLQFNVGAREGLADADVLHVDMAGLTSFAG-----GHRLL 449
Qy      769 TGGIAGYIMSNFNHVPYRLRFEPAGDOSIRGYAHDLSPISDGYLTGGVLAVGTA 848
Db      450 IGRLOVGIATNDYKSTPSLRFEPAGDOSVGRDYRTISPENSODKIGGKRYMAGSV 509
Qy      849 YNFERMKDLRLAVFGDIGNAYDKFTNDTKIGAGVGRMASPVGOVRVDVATGVKKEGMP 908
Db      510 YQYPLAEHRRLATFVDQGNAFNSLDPSIKTGVGGRVWSPVGPLRLDLAALADDG 568
Qy      909 IKLHFEIG 916
Db      569 FRLHFSMG 576

RESULT 2
O9CK26 PRELIMINARY: PRT: 586 AA.
AC O9CK26:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein PM1809.
GN PM1809.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PM70;
MEDLINE=21145866; PubMed=11248100;
RA May B.-J., Zhang Q., Li L.-L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida PM70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL: AE006218: AAK03893.1: -;
GN InterPro: IPR001084; Bac_surfa9-D15.
DR InterPro: IPR001092; H4H_basic.
DR Pfam: PF01103; Bac_surfa9_Ag_1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
```

```
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 586 AA; 67106 MW; 00895174DDE58283 CRC64;
Query Match 11.3%; Score 532; DB 16; Length 586;
Best Local Similarity 23.0%; Pred. No.3.8e-23;
Matches 166; Conservative 107; Mismatches 258; Indels 192; Gaps 14;

Qy 204 OKTEPYANIKAALEDITQESANDLNGSIPRLKOTALVAARAVGY--DIDLSIINSGE 261
Db 41 QNTQNNNVRIYVGMIDKEADGSEHKKQLYREALDKLARAGYQSEVEFQIESQPK 100
Qy 262 VD-VIIH-DLGEPIYIDRAVEYRGEADDKAFTTYVADEYFLLIDGVPHGKYETKNNI 319
Db 101 KQDLIAHKYIKGPKIADYDEVYLGATODPEFIATVKIIPK-CSIINHETDYNKSHI 159
Qy 320 ENASAEHGYFDGRWILDRSDVILIPNTADVSLIYDTGTQYRDEYVFTIDPKTNQLTDT 379
Db 160 QKIALSRGYFDGDFLVSLREVRPSTQQAAMWRLDYDSGEYRGEYTF----- 206
Qy 380 PDKLPVKRELLECLLTVMNGEAYNLQAVRALSNDLIATRYFNNVNTETIVEPEREQINDQ 439
Db 207 -ENAOIRREDYLNMMINFNKGOPYLLINDSTLTNNYSSNMFSSVLMQPYLDEHKIVND 265
Qy 440 VFEEOSSSRTPEAQNDESTLEPIETVELTDGILMDISPIESASNLIDDKLNLVAAKA 499
Db 266 V-----LLQPR----- 271
Qy 500 RHLVMPDRVLA1NHDDGVNRSILGRISDAVASAVARAILPDESENEVIDLPERTALANR 559
Db 272 ----- 271
Qy 560 KTPADVYQSKKVPVLYFVASDKPRDQIGLGWGSDTGTRLVTKFEHNLINRDGYQAGEL 619
Db 272 -----KKNSMEVIGMASDGPRLQGLWTKPWLNNGHSFRNL 310
Qy 620 RLSEDKKGVK-LYATKPLSHPLNDLRLATLGQOEVEFHSINGPDLSTRTLHEISR 675
Db 311 YVSARKQTLFATYKMPMLKNPLNTYYEYAGLENE-NKNDTSSPASLSAIAYWNHE--- 366
Qy 676 STIONGMNRTYSLRYRLDKLTKQAPEPTWQDLPVDFVNGKSGOELLN--GVAAHKYIA 733
Db 367 -----AGWGHSHGLRYRYD-----SFIDANKDKTLLEPTTASVARTRL 405
Qy 734 DNLVNPFRGYRQKRSLEVGSSGLVSDANMALIARAGISGVYSGFNATVGSRAHOMGTIO 793
Db 406 QCGLEPTTWCDTOKLTITDLRTWMLSDVDLKMGGSSLMVRYTLQH-----HRTYRL 458
Qy 794 AGYIMSDNFNHVPYRLRFEPAGDOSIRGYAHDLSPISDGYLTGGVLAVGTAEYNEF 853
Db 459 LGMLHTKNIERILPPALRFEPAGDGRSIRGYKKIAPKNAGKLVGSRLLTGSFEYQYQV 518
Qy 854 MKDLRLAVFGDIGNAYDKFTNDTKIGAGVGRMASPVGOVRVDVATGVKKEGMPKILHF 913
Db 519 YPDWMLATPADTGLANQFTTELRYGAGMGVWSPVGAIKFIDLATTPVDRDNSKNIOF 578
Qy 914 FIG 916
Db 579 YTG 581

RESULT 3
O8ZB95 PRELIMINARY: PRT: 578 AA.
AC O8ZB95:
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative exported protein.
GN YP03524.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
```

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CO-92 / BIOVAR ORIENTALIS;
 RX MEDLINE-21470413; PubMed-NR;
 RA Parhill J., Wren B.W., Thomson N.R.,
 RA Prentice M.B., Sebaiha M., James K.D., Churchill C., Houlden M.T.G.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Felwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
 RA "Genome sequence of Yersinia pestis, the causative agent of plague";
 RL Nature 413:523-527(2001).
 DR EMBL: AJ414157; CAC92753.1; -
 DR InterPro: IPR000566; Lipoclin_cytfabp.
 DR PROSITE: PS00213; LIPOCALIN: UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 578 AA; 64405 MW; 69AD3780BFC20F1 CRC64;

 Query Match 10.5%; Score 494; DB 16; Length 578;
 Best Local Similarity 23.4%; Pred. No. 6.5e-21;
 Matches 167; Conservative 105; Mismatches 255; Indels 188; Gaps 17;

 QY 211 NIKAALEDI-TOSAMDLNLSI-PLRQRTALVAAKANGYD--IDLSTIKNSIGEVDYII 266
 DB 38 NVAAKSLTGTDEVTAD--GRFRSRVDEALRQGLRALGYDPTTTELRQRRPAPARSVLI 95
 QY 267 HDL--GEPYIDYRAVEVREGADKFAFTVADEVPLLDGVFHGKYEKKNLINENASA 324
 DB 96 AKVYBEPVILAGVDVLOGAKTDPDYQALVRRDTPKISSVLNKGDFDFTSSLTGLAL 155
 QY 325 EHGCFDGRWLDKRSVDYILPNTADVSLIYDTGYRDEVEVFTIDPKTNQLTTPDKLP 384
 DB 156 RRCYFDANMIKSQLGVAQAQHEAFWMDIDFDSGGRGKVIK-----QGSQ 201
 QY 385 VKELLEQLLTVMKGEYNNQAVRAISNDLIATRYFMVNTETVYFPRRQIQNDQVFEQ 444
 DB 202 IRDYLQNLVLPFHEGERTYSDELAELNRRLLAATNMFNSV--VVSPP-----FOD 248
 QY 445 SSSRTEPAQVDSLEPIETVELTDLGILMDSPIEFSSNLIDQKLNIAAKARHLVD 504
 DB 249 AKRSK-----LPLD----- 258
 QY 505 MPDRLVAINHDDGVNKSILGRISDAVASARAAILPDESENEVIDLPERALANRKTAD 564
 DB 259 -----AVVPRPENTV----- 269
 QY 565 VYOSKAVPLVYASDKPRDQGLGWSGTGRLVTKFEHNLINRDGYQAGAEIRLSED 624
 DB 270 -----ELGGYATDVGPRLTASMKRPMWNSFGHSLTTTALSAP 308
 QY 625 KKGVL-YATKPLSHPLNDQLRATLGYQDEVFGHSTGFLSPFTLEHLSRISLIQNGW 683
 DB 309 EQGLDSTYRPLRNPLEQYLLIQQGFR-----IDLNDTNSDTTLLNVAREWLSGGW 362
 QY 684 NRTYSRLRYLDLKTQAPRTWDLPYDFVNGKPSQAL--AGVAHKTVAIDLVPNR 741
 DB 363 QRAINLRMSLD-----HFTQGRVTDFTMLLPYGVSLNTRRQRCGAPRW 406
 QY 742 GYQRIISLEVSSGLVSDANMAIARAGISGVSEFGDNAYGNSNRHQTGCIQAGYIWSDN 801
 DB 407 GDSQRSIDVSDTTCGSDVDGFIQONWVIRLGEK-----NREVARGNWMIETNN 459
 QY 802 FNVVPRRLRPFAGGDSIRGYANDSLPIDKGYLLGGOVLAVGTAVYENPMKDLRLAV 861
 DB 460 FDEVPSSLRFAGGDSIRKTKRNTSPRDSGKLTGASGLATGSLLEYQYNFTGRMGAV 519
 QY 862 FGDIGNAYDKGFTNDTKIGAGVGVNRASPVGYRVADVATGVKKEGNDIKLHFFIG 916
 DB 520 FVDSGAVNNFSSDKLTGAGVGVNRASPVGPKIKLDIAAI--GDNEHNGVQFYIG 573

 RESULT 4

Q8XELS
 ID Q8XELS PRELIMINARY; PRT; 577 AA.
 AC Q8XELS;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 GN YTFM OR STM4409 OR STY4768.
 OS Salmonella typhimurium, and
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_taxid=602, 601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX SPECIES-S.typhimurium; STRAIN-LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE-21534948; PubMed-11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2";
 RL Nature 413:852-856(2001).
 RT [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RX SPECIES-S.typhi; STRAIN-C718;
 RX MEDLINE-21534947; PubMed-11677608;
 RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Houlden M.T.G., Sebaiha M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gara P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RA "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi C718";
 RL Nature 413:848-852(2001).
 RX EMBL: AE008906; AAL23329.1; -
 DR EMBL: AL627283; CAD06889.1; -
 DR InterPro: IPR000184; Bac_surfAg_D15.
 DR InterPro: IPR000566; Lipoclin_cytfabp.
 DR Pfam: PF01103; Bac_surface_Ag_1.
 DR PROSITE: PS00213; LIPOCALIN: UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 577 AA; 64765 MW; FBAA190173BFL8C CRC64;

 Query Match 10.1%; Score 479; DB 16; Length 577;
 Best Local Similarity 22.7%; Pred. No. 5e-20;
 Matches 166; Conservative 104; Mismatches 257; Indels 204; Gaps 17;

 QY 210 ANIKAALEDITOSAMDLNLSIPLRQRTALVAAKANGYD--IDLSTIKNSIGEVDYII 262
 DB 22 ANVAAKSLTGTDEVTAD--GRFRSRVDEALRQGLRALGYDPTTTELRQRRPAPARSVLI 95
 QY 267 HDL--GEPYIDYRAVEVREGADKFAFTVADEVPLLDGVFHGKYEKKNLINENASA 324
 DB 96 AKVYBEPVILAGVDVLOGAKTDPDYQALVRRDTPKISSVLNKGDFDFTSSLTGLAL 155
 QY 325 EHGCFDGRWLDKRSVDYILPNTADVSLIYDTGYRDEVEVFTIDPKTNQLTTPDKLP 384
 DB 156 RRCYFDANMIKSQLGVAQAQHEAFWMDIDFDSGGRGKVIK-----QGSQ 201
 QY 385 VKELLEQLLTVMKGEYNNQAVRAISNDLIATRYFMVNTETVYFPRRQIQNDQVFEQ 444
 DB 202 IRDYLQNLVLPFHEGERTYSDELAELNRRLLAATNMFNSV--VVSPP-----FOD 248
 QY 445 SSSRTEPAQVDSLEPIETVELTDLGILMDSPIEFSSNLIDQKLNIAAKARHLVD 504
 DB 249 AKRSK-----LPLD----- 258
 QY 505 MPDRLVAINHDDGVNKSILGRISDAVASARAAILPDESENEVIDLPERALANRKTAD 564
 DB 259 -----AVVPRPENTV----- 269
 QY 565 VYOSKAVPLVYASDKPRDQGLGWSGTGRLVTKFEHNLINRDGYQAGAEIRLSED 624
 DB 270 -----ELGGYATDVGPRLTASMKRPMWNSFGHSLTTTALSAP 308
 QY 625 KKGVL-YATKPLSHPLNDQLRATLGYQDEVFGHSTGFLSPFTLEHLSRISLIQNGW 683
 DB 309 EQGLDSTYRPLRNPLEQYLLIQQGFR-----IDLNDTNSDTTLLNVAREWLSGGW 362
 QY 684 NRTYSRLRYLDLKTQAPRTWDLPYDFVNGKPSQAL--AGVAHKTVAIDLVPNR 741
 DB 363 QRAINLRMSLD-----HFTQGRVTDFTMLLPYGVSLNTRRQRCGAPRW 406
 QY 742 GYQRIISLEVSSGLVSDANMAIARAGISGVSEFGDNAYGNSNRHQTGCIQAGYIWSDN 801
 DB 407 GDSQRSIDVSDTTCGSDVDGFIQONWVIRLGEK-----NREVARGNWMIETNN 459
 QY 802 FNVVPRRLRPFAGGDSIRGYANDSLPIDKGYLLGGOVLAVGTAVYENPMKDLRLAV 861
 DB 460 FDEVPSSLRFAGGDSIRKTKRNTSPRDSGKLTGASGLATGSLLEYQYNFTGRMGAV 519
 QY 862 FGDIGNAYDKGFTNDTKIGAGVGVNRASPVGYRVADVATGVKKEGNDIKLHFFIG 916
 DB 520 FVDSGAVNNFSSDKLTGAGVGVNRASPVGPKIKLDIAAI--GDNEHNGVQFYIG 573

 RESULT 4

```

Oy 491 KLNLYAAKARHLYPDMPDRVLAINHDDGVNRSILGRISDAVSAVARAILPDESENEVIDL 550
Db 256 -----LKGVSPRETEIT----- 268
Oy 551 PERTALANKRTPADYQSKVPLVYFVASDKPRDQIGLWGSDDTGRILVTKFEHNILNR 610
Db 269 -----ETGVGSTVDGPRVKASMKPMNNS 293
Oy 611 DGYOAGELRLSEDKKGVKL-YATKPLSHPLNDQRLATLGVQEVFGHSTGFDLSTRTL 669
Db 294 YGHSLLTSTISABEQVLEDFSTKMPKLKNPLQYLYLVQGGFKRT---DLNDTDDSTTL 349
Oy 670 EHEISRSIIIONGNRRYSLRYRLDKLTKQAPPETWDLPPVFNKPSQELAL--AGVA 727
Db 350 --AVSRYWDLSSGQRAINLRWSPD-----HFGQNVNTNTMLFYECVM 391
Oy 728 VHKTVANLVNPMKGYQORISLEVSGSLVSDANMAIARAGISGVYFSGDNAVGSNRAHQ 787
Db 392 ISRTSRKGLMPTWDSQORSVDYSNTAMGSDYDFVLQANWMIRTLDR-----HR 444
Oy 788 MTGGIAGYIWSDNFNHVPRLRPFAGDOSIRGVAHDSISPSIDSKGYLGGVLAAGTA 847
Db 445 FVMRANLGMLETODFDPKVPPLDRPFAGDGRSIRKTKTSISPKSDGNLKGASKLATGSL 504
Oy 848 EYNEFMKDLRLAVFGDIGNAYDKGFTNDTKIGAGVGRVAPGVQYRVQ--VATGVKEE 905
Db 505 EYOYNTYTGKMGAVFVDSGSAVSDIRSDPKTGTGVGVNRASPVGPKLDPFAPVGDKE 564
Oy 906 GNPYKHLFFIG 916
Db 565 HG---LQFYIG 572

```

```

RESULT 5
O9KP31 PRELIMINARY: PRT: 582 AA.
AC O9KP31:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Hypothetical protein VC2548.
GN VC2548
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RT Nature 406:477-483(2000).
RL EMBL: A0004323; AAF95689.1; -.
DR TIGR: VC2548; -.
DR InterPro: IPR000184; Bac_surfaG_D15.
DR InterPro: IPR000566; Lipocin_cytfABP.
DR Pfam: PF01103; Bac_surfaG_Ag_1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 582 AA: 65644 MW: 696946FB85F1C426 CRC64;

```

```

Query Match 9.6%; Score 454; DB 16; Length 582;
Best Local Similarity 22.4%; Pred. No. 1,5e-18;
Matches 164; Conservative 104; Mismatches 237; Indels 228; Gaps 20;
Oy 211 NIKAALEDITQESAMDLNGSI---PLRQTALVAAVAVY--DIDSLIRNS----- 258

```

```

Db 48 NVEAYLSI---AADYSTSLRFQSLERSMTEALNALGYHPSIDFTVSEDNQLRAAV 104
Oy 259 -----IGEVNVIITHDGEPEYIDYRAVEVGBEADKAFITVADENVPLIGVFHHGK 311
Db 105 TLGEVTRLSEVDIVL-----RGEAGGRDQRLRRSGRLVADAPLNSL 148
Oy 312 YETKNLLENASAEHGYEDGRWLDKRSVDVILPDNTADVSLIYDGTGYRPFDEVFETIDP 371
Db 149 YDNLSKSGIRNLALQGYRNGDFQASRLLEVIPELNARVILHFDGIRLFL----- 198
Oy 372 KTNQLTTPDKLPVKRELLEOLLTVNMGAYNLQAVRALSDLLATRYFNMYNTEIYFPE 431
Db 199 --GATTEGSGIDENRW--SLRPFKQGEPLYVQVGFQNLSTDMFSSVPE--PD 251
Oy 432 REQIONDVFSFGSSSSSTPEPAQVDESLREVIETVELTDLMDISPIERSASNLIDOK 491
Db 252 LSQI----- 255
Oy 492 LNLVAARHLYPDMPDRVLAINHDDGVNRSILGRISDAVSAVARAILPDESENEVIDLP 551
Db 256 -----DEG----- 258
Oy 552 ERTALANKRTPADYQSKVPLVYFVASDKPRDQIGLWGSDDTGRILVTKFEHNILNRD 611
Db 259 -----RELPIKVTTLAPQARNQLETGIGSTVDGVGSLKMKKPMWNSQ 301
Oy 612 GYOAGELRLSEDKKGVKLATKPLSHPLNDQRLATLGVQEVFGHSTNGFD-LSTRTL 670
Db 302 GHSFSSFSLSIPEQTTTAACTKIPLEDALNEYRYIOYGMK-----LDRDRTSLE 352
Oy 671 HEIS--RSIIIONGNRRYSLRYRLDKLTKQAPPETWDLPPVFNKPSQELLAGVAV 728
Db 353 SNLSLEHHQDLDGWMHRTVIRLLENYR-----QGIQDD-----NSQFLLRPMY 398
Oy 729 HKT-VADNLVNPARGYRQORYSLEVSGSLVSDANMAIARAGISGVYFSGDNAVGSNRAHQ 787
Db 399 TRTRTSNSGLTLWGDQKQTTLEYGDPALLSETRVRLQTGSSWMTYRAN-----HR 451
Oy 788 MTGGIAGYIWSDNFNHVPRLRPFAGDOSIRGVAHDSISPSIDSKGYLGGVLAAGTA 847
Db 452 ALVRVYDGANLVDEFQQLSPSLRPFAGGNLNGRYKYSIPQDASGALTGAAYIATSS 511
Oy 848 EYNEFMKDLRLAVFGDIGNAYDKGFTNDT---KIGAGVGRVAPGVQYRVAVGV-K 903
Db 512 EYOYRLTGMWMAAMFMDVGDAP-----NPNPKKGVGTGIRWISFVGPIRLDFAWGLDA 566
Oy 904 EGNPKHLFFIG 916
Db 567 APGDEKIHFFIG 579

```

```

RESULT 6
O9EY32 PRELIMINARY: PRT: 593 AA.
AC O9EY32:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Putative outer membrane protein.
OS Xanthomonas oryzae pv. oryzae.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX Xanthomonas.
OX NCBI_TaxID=64187;
RN [1]
RP SEQUENCE FROM N.A.
RA Goel A.K., Rajagopal L., Nagesh N., Sonti R.V.;
RT "Xanthomonadin biosynthetic gene cluster from Xanthomonas oryzae pv.
RT oryzae."
RT Submitted (MUG-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AY010120; AAC38834.1; -.
SQ SEQUENCE 593 AA: 66684 MW: 0FF8A8C135C4F91A CRC64;

```


Db 473 LRWHGAGASSRLIRG-----EAGTTWASDLVAMPSPSLREFAGVNSIRGVAFREVG 525
 QY 829 PISDK--GYLTGGVAVTAETAEYNEFM-KDLRLAVFGDIGNAYDKGTNTPTKIGAGVY 885
 Db 526 PRTAKPDAFALGAKHVFSAGAEFEYFYKGGPFGGAVFVDSASAFNR--YRPMHTGVGIGL 583
 QY 886 RMASPVGOVRDVATGKVEEGNPILKHEFFIG 916
 Db 584 KYRSPGVPVRDIARGINSPSDKIOLYIDIG 614

RESULT 8

OBRRHT4 PRELIMINARY: PRT: 678 AA.
 ID OBRRHT4
 AC OBRRHT4
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Outer membrane protein.
 GN FN1911.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteria; Fusobacterium.
 OX NCBI_TaxID=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RX MEDLINE=21886394; PubMed=11889109;
 RA Kapatala V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
 RA Larstein M., Kyriades N., Malunas T., Pusch G., Haselkorn R.,
 RA Fongstein M., Kyrpides N., Overbeek R.;
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium
 RT nucleatum strain ATCC 25586.";
 RL J. Bacteriol. 184:2005-2018(2002).
 DR EMBL: AE010493; AAL94010.1;
 KM Complete proteome.
 SQ SEQUENCE 678 AA; 76229 MW; FB3579AFD639A74B CRC64;

Query Match 5.6%; Score 262.5; DB 16; Length 678;
 Best Local Similarity 22.6%; Pred. No. 4e-07;
 Matches 162; Conservative 111; Mismatches 265; Indels 179; Gaps 36;

QY 311 KYETKKNLIE-NASAEHGFDGRMLDRSDVILPDNTADVSLYDTGTQYRFD---EYV 365
 Db 33 KESTLEALLDFNLKKEGYFE-----DVILQ-----PVSVDGVRIVDVVEKENV 78
 QY 366 FTIIDPK---TNGLTDPKLPY-----KRELLEOLLIVNM--GEAYNLQAVRAL 410
 Db 79 VDLKKEGVAINTLRDQDSIVLSVKTGMKRVTTSELLDITOLKACGYFSRVEDA 138
 QY 411 SNDLIATRYFNMYTEIVFEREQIONDQ--VSFEQSSSRTEPAQVDESLTLEIETVE 468
 Db 139 QRRLLATGKFSVEVRPD-----AQVANGKMAISFE-----VVEN--PIKSVI 178
 QY 469 LFDGILMDISPI--EESA-----SNLIQDK--LNLVAAKARHLQYMPD-----DRV 510
 Db 179 ITCNNNTIPTSTIMSELTTKPGSVQYNNLRDRDKILGLYQAGYTLVNTIDKSTDENGT 238
 QY 511 LAIINHDDGVNRSILGR-----ISDAVASAARAILPDESE---NEVIDLPERTA 555
 Db 239 LHISTVEGIVRIEYKKNVTKOKGNRRTPNDVLKTKDVYIDREIIOPIKIFNVKEYDA 298
 QY 556 LANRRTPADV-----YQSKVP-----LYYFVASDKPRDQIGLQWGSQDGTSLVTKF 603
 Db 299 TVDNLMLRGLIFKNVKEYEASIFQDPEGIDILILIDEDKRAELQAGVAVSSEFG-----F 352
 QY 604 ENNLINRCGYQAGAE---LRLSEDKKGVKLYATKPLSHPLDQLERATLGYOQEVFGHST 659
 Db 353 LCTLSLKDSNMWKGKQOQFGFTFEKSNKNTYGFALDFYDPMKIDTDRVSGW-----GAYR 407
 QY 660 NGFDLSTRLHEIS---RSIIQNG-GNNRRTYSLRYRLDKLTKTQ----- 699

Db 408 TSYGDESDILFHEIDTIGFRTNIGKGLGNFTLSLGTQVEYIEKEHDEQKLRQANNGKMY 467
 QY 700 -APPTMODLPVDFPNGKPSQEALLAGVAVKTVADNLVNPMPKGRKORXSLEFVSSGLVS 758
 Db 468 YKEKKMKRE--IEGVDDKYLWMSITPIYIS--DTNNNNLNPSTSGFYGFQVEAGHAGCY 523
 QY 759 DANMAIARAGISGYVS--FGDNAYGSNRHAWMTGTGIGQAGYIMSDNFNVPRVPLRFAGD 816
 Db 524 SGNCGNATLELRTYKIKGLFKNNIF-----AKYVGGAIV-----NNTKESQKFWGG 571
 QY 817 QSRGVAHDSLSPISSDKGYLTGGVAVTAETAEYNEFMKDLRLAVFGDIGNAY-----DK 871
 Db 572 NSLRGY-----DGGEFFKGSQKL-VATIENTRQLNDIIGLYFADAGRAMKNGRDP 621
 QY 872 GFTNDT-----KIG--AGVGYRAMSPVGOVRVDA--TGKVEEGNPILKHEFFIGTPF 919
 Db 622 SYTRDNSRFNGHIGTGTAGVGIRLNTPIPIGLRFDGWPVGNKMDDMGKMFYFMGGSF 678

RESULT 9

P95359 PRELIMINARY: PRT: 792 AA.
 ID P95359
 AC P95359;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Outer membrane protein.
 GN OMP85.
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=485;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FA19;
 RA Manning D.S., Reschke D.K., Judd R.C.;
 RT "OMP85 proteins of Neisseria gonorrhoeae and Neisseria meningitidis
 RT are similar to Haemophilus influenzae D-15-Ag and Pasteurella
 RT multocida Oma87.";
 RL Microb. Pathog. 23:0-0(1998).
 DR EMBL: U01959; AAC17600.1;
 DR InterPro: IPR000184; Bac_surfAg_D15.
 DR Pfam: PF01103; Bac_surface_Ag_1.
 SQ SEQUENCE 792 AA; 87868 MW; 90E32D24AA0513D8 CRC64;

Query Match 5.5%; Score 259.5; DB 2; Length 792;
 Best Local Similarity 20.6%; Pred. No. 7.9e-07;
 Matches 189; Conservative 119; Mismatches 355; Indels 255; Gaps 37;

QY 116 TPLSLELFAQESTEMGINDND-----YIPEYQGEQPNSEVVPPTLPEKEKQLIKRLY 169
 Db 16 SPLAFADFTIQDIRVEGLQRTPESTYFNLYPVKVGDTYN-----DTRGSAIIRSLY 66
 QY 170 ARLENDGV-----NKPRLKAKFYQSSQ--SGETSAGISSHOKTEPY 209
 Db 67 ATGFEFDDVRVETADGILLLVIVCPTIGSLNITGAKMLQNDALKKNLESGLQASQVFNQ 126
 QY 210 ANIKALEDTQSSADNLNSITRLKQTFALYAARANGYIDISITRNSIGEVDVYIIDL 269
 Db 127 ATLNOAVAGKEEY---LGRKLNLIOTTPKVTKLARNRVDIDITIEGKSAKITDIEFE- 182
 QY 270 GEPVYIDYRAVEVREGGADDAKFTVADEVPILIGVFNHKKYETKKNLLENMSAEHGYP 329
 Db 183 GNGVYSDKILMR-----QMSLTBEGIGTWLSTRSDRFDRQKFAQDMEKYTDYQNNNGYF 235
 QY 330 DGRWLDRSVDVILPDNTADVSLIYDTGTQYRFDDEVVFTTIDPKTNOLTTDDPKLPVKREL 389
 Db 236 DFRILDTDIQTNEDKTRQRTIKITVHEGGRPMGKV---SIEGDTNE-----YPKAE 283
 QY 390 LQGLTLVNMGEAYNLQAVALSNDLIATRYFNMYNTEIVFEREQIONDQVSFEQSSSR 449
 Db 284 LEKLTLTKPKGKYERQOMTAVILG-----ETONMGSGAGVAYS-- 320
 QY 450 TEPAGVDESTLEPVIETVELTQGIIMDISP-----IEFSASNLIQKLNLYAAKAKAH 501

```
Db 321 -----EISVQPLPNAGTGTVDVFLHIEFGKRIYVNEHIITGNKTRDE--VVRRLRQ 371
Oy 502 LYDMPD-----RYLAINHDDGVNRSILGRISDAVSAVARAILPDESENEVIDLPE 552
Db 372 MESAPYDTSKLORSKERVELLGFEDNVQ-----FPAVPLAGFPDK-----VDL-- 414
Oy 553 RTALARKRPADVYQSKKVLVYFVASDKPRDQIGL--GMSDPTGRVLTKEEHLINR 610
Db 415 NMSLTERST-----GSLDLSAGMVQDTG--LVMSAGVSQDNL 449
Oy 611 DGVOAGAEILRSEDKGVKLYATKPLSHPLNDQLRATLGQQVEFGSTGFDLSTTLE 670
Db 450 FGTGKSAALRASRK--TTLNGSLSTDPYFTADGVSLG--DIYGAFFPKRASTSVKQ 505
Oy 671 HEISRSIIIONGWNRTYSLYRLDKLTQAPPE-----TWODLP--VDFVN--GK-PSQ 719
Db 506 YKTTTA-----GGGVRMGIPRTYEDRVNFGLAHLVNTYNKAPKRADIKQYKTDGA 561
Oy 720 EALLAGVAAHKTY-----ADNLVNPARGYRQRYSLIEVGSSGLVSDANMAIARAGISGVY 773
Db 562 DGSFKGLYKGYVGMGRNKTDLSALMPTRGY-----LTGVNAEIALPGSKLYQ 608
Oy 774 SFGDNAYGSRBAHQMG-----GIOAGYIWSDNFNHVPRLRFPFAGDQ 817
Db 609 -----YSATHNOTWFPPLSKFTTLMLGEGVGIAGY--GRTKEIRPEFENEYGGGLG 657
Oy 818 SIRGVANDLSPIIDSKGYLTGQVLAVG--TAEYNYEFM-----KDLRLAVEGDI 865
Db 658 SVRGYSSTGLCP--KVYDEYGEKISYGGKKNKAVNSALLEFPMFGAKADATVRLSLFDA 714
Oy 866 G-----NAVDKGFTNDTKIGAGVGVMASSPVGQVADVATGVK 903
Db 715 GSWMDRTYTAENGNNKSVYSENAHSTFTNELRYSAGAVTWLSPLGPMKFTYAVPLK 774
Oy 904 E--EGNPRIKHFPIGMPF 919
Db 775 KKPEDEIQRFQOLGTFE 792

RESULT 10
O9K1H0 PRELIMINARY: PRT: 797 AA.
AC O9K1H0.
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Outer membrane protein Omp85.
GN NMB0182.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCB1_TextID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGRUP B;
RX MEDLINE=2017575; PubMed=10710307;
RA Tetteiln H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Hatt O.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Collon M.D., Ulteback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gull J., Scarlato V., Masiugnani V., Pizsa M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58".
RL Science, 287:1809-1815(2000).
DR EMBL: AE002375; AAF40639.1; -.
DR TIGR: NMB0182; -.
DR InterPro: IPR000184; Bac_surfaG_D15.
DR Pfam: PF01103; Bac_surface_Ag_1.
KW Complete proteome.
SO SEQUENCE 797 Aa; 88436 Mw; B68BFC5A3D22EE8 CRC64;
```

```
Query Match 5.5%; Score 259; DB 16; Length 797;
Best Local Similarity 20.5%; Pred. No. 8.6e-07;
Matches 191; Conservative 121; Mismatches 342; Indels 278; Gaps 40;

Oy 116 TPISLEELFQSESTEMKINND-----YIPYQGEOPNSEVVPVPLLEPEKELIRLY 169
Db 16 SPLADFTIÖDIRVELOQTEPSTVENYLPVKGDYNN-----DRHSAIIISLY 66
Oy 170 ARLEFNDGVNVP--RLKAFYQSSQSEMSAIGSSHQKTEPVANIKALED--ITQESA 224
Db 67 ATGFEDVRYETADGÖLLLVIERPTGISLNTGAKKLÖDA--IKNLSEFGLAQSQY 123
Oy 225 MD--LNGSIPLRQ-----TALVAARAVGYIDLSIIRNSIGEVDIITDL 269
Db 124 FNQATLNQAAVAGLKEEVLGKGLNIQITPVYTKLARNRVIDIDITDGSKAKITIDIEF- 182
Oy 270 GEPLYIDYRAVEYRGEGADKAFITVADEVPLIGDV-----FHNGYEFKKLIE 320
Db 183 GNOVYSDRKLMR-----QMSLTEGGIWTWLTNSNQFNEQKFAQDMKEVT 226
Oy 321 NASAEHGFEDGRMLDRSDVILPDNTADVSLIYDTGYQFEDEVFTTIDPKTNQLTDP 380
Db 227 DFTÖNNGFEDFRILDDIÖTNEDKTQÖTITVHEGGFRFMGV--SIEGDYNE----- 278
Oy 381 DKLPVKRELLEÖLLTVNMGEAYNLQAVRALSNDLIATRYFNMVTEIYFPEREQIQDQV 440
Db 279 --VPRAELEKLLTMKRGKMYERQÖMTAVLG-----EIQRMG 313
Oy 441 SFEÖSSSSTREPAQVDESTLEPIVETVELTÖGILMDISP-----IEFSASNLQÖDKL 492
Db 314 SAGYAVS-----EISVQPLPNAETKTVDVFLHIEFGKRIYVNEHIITGNKTRDE- 363
Oy 493 NLVAAKARHLYDMPD-----RYLAINHDDGVNRSILGRISDAVSAVARAILPDES 543
Db 364 -VVRRLRQESAPYDTSKLORSKERVELLGFEDNVQ-----FPAVPLAGFPDK- 411
Oy 544 ENEVIDLPERTALANRTPADVYQSKKVLVYFVASDKPRDQIGL--GMSDPTGRVLT 601
Db 412 ----VDL--NMSLTERST-----GSLDLSAGMVQDTG--LVN 440
Oy 602 KFEHNLINRDGYOAGAEILRSEDKGVKLYATKPLSHPLNDQLRATLGQQVEFGSTNG 661
Db 441 SAGVSODNLEGTGKSALRASRSK--TTLNGSLSTDPYFTADGVSLG--DYYGKAFDP 496
Oy 662 FDLSTRLHEIRSIÖNGGNR-----YSLRYRLDKLTQAPPTWODLP--VDFVN 713
Db 497 RKAÖTSIKÖYKTTTA--GAGIRMSVPTYEDRVNFGLAHLVNTYNKAKRKHVADFT 552
Oy 714 N--GK-PSÖALLAGVAAHKTY-----ADNLVNPARGYRQRYSLIEVGSSGLVSDANMAI 764
Db 553 KKYGKTDGIDGSKFKMLYKCTVGMGRNKTDLSALMPTRGY-----LTGVNAEI 599
Oy 765 ARAGISGVYSFGDNAYGSRBAHQMG-----GIOAGYIWSDNFNHVPYR 808
Db 600 ALPGSKLYQ-----YSATHNOTWFPPLSKFTTLMLGEGVGIAGY--GRTKEIRPF 648
Oy 809 LRFPAQGDÖSIRYANDLSPIIDSKGYLTGQVLAVG--TAEYNYEFM-----KD 866
Db 649 ENYFGGGLGSVRCESGTIGP--KVYDEYGEKISYGGKKNKAVNSALLEFPMFGAKDART 705
Oy 857 LRLAVEDIGNAVD-----KÖFTNDTKIGAGVGVMASSPVGQVADVATGVKMS 889
Db 706 VRLSLPADGGSVWDGTYTDNNSATGKRVÖNTYGCNTHKSTFTNELRYSAGAVTWLS 765
Oy 890 PVGÖVRVDVATGVKE--EGNPRIKHFPIGMPF 919
Db 766 PLGPMKFYAYVPLKKKPEDEIQRFQOLGTFE 797

RESULT 11
O9UXJ1 PRELIMINARY: PRT: 797 AA.
```


AC 09JX31;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Outer membrane protein OMP85.
 OS OMP85 OR NMA0085.
 GN Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=56599;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=2491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S.,
 RA Jagers K., Leach S., Moule S., Kung'u K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrall B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis 2491."
 RL Nature 404:502-506(2000).
 DR EMBL: AL162752; CAB83401.1;
 DR InterPro: IPR00184; Bac_surfAg_D15.
 DR Pfam: PF01103; Bac_surface_Ag; 1.
 DR Complete proteome.
 KW SEQUENCE 797 AA; 88404 MW; 65DE47E0C9E1D1F CRC64;

Query Match 5.5%; Score 259; DB 16; Length 797;
 Best Local Similarity 20.5%; Pred. No. 8.6e-07;
 Matches 191; Conservative 121; Mismatches 342; Indels 278; Gaps 40;

OY 116 TPLSELEFAQESTENGINPND-----YIPEVQGEOPSEVVVPTLEPEKGLIKRLY 169
 DB 16 SPLALADFLIDIRVGLQRTSTPYNILPVAVGDTYN-----DHGSAILIKLY 66
 OY 170 ARLFNDGVKVP---RLAKKPYOSSGSETSAIGSSHOKTEPYANKALED--ITQESA 224
 DB 67 AGCFEDVDVETADGQLLTIVIERPTIGSLNITGAKMLQND---IKNLESGLAQSOY 123
 OY 225 MD---LNGSIPRLRO-----TALVAARAVGYDIDISTIRNSIGEVDYIHD 269
 DB 124 FNOATLNOAVAGLEKEYLGRGLNIOITPKYKLARNRVIDITIDEKGSAMTIDIEFE- 182
 OY 270 GEPYIDYRAVEVRGEGADKAFITVADEVPLIGDY-----FHGKXETKKNLE 320
 DB 183 GNGVYSDRKLAM-----QMSLTGEGIMTWTLSNOFQKFAQDMEXYT 226
 OY 321 NASAEHGFDCGMLDRSDVILPDTADVSLIYDTGYRDEVVFFITDPTNQLTDP 380
 DB 227 DFYQNGGYFDFRILDTDIQTNEDKTKOTIKIVHGGFRMKV---STEGDTNE----- 278
 OY 361 DKLPRKRELLBOLLVNNGEAIVLQAVRALSNDLIATRYFNWNTIYFPEREQIONOV 440
 DB 279 ----VPKAELELLTMKKGKWERQOMTAVLG-----EIONRMG 313
 OY 441 SFEQSSSRTEPAQVDESTLEPIETVELTDLGILMDISP-----IEFSASNLIOK 492
 DB 314 SAGAVYS-----ELSVQPLPAETKTYDVFVLAHIEGRKIYNEIHTGNNTKTRDE- 363
 OY 493 NLVAAKARHLTYDMPD-----RVLAINHDDGVNRSILGRISDAVASAVARALLPES 543
 DB 364 -VVRRELROMESAPYDTSKLQRSKERVELLYFPDVQ-----PDVAPLAGTPEK- 411
 OY 544 ENEVYIDLPERTALNKRTPADYOSKKVPLVFAASDKPRDQIGL--GWGSDGTGRLVT 601
 DB 412 ----VDL--NMSLTERST-----GSLDLSAGWVDTC--LVM 440
 OY 602 KFEHNLIRPDYQAGAEILISEDKKGVLYATKPLSHPLNQLRATIGYQOEYGHSHNG 661
 DB 441 SAGYSQNLFCOTGSAALRASRSK--TTLNGSLSTFDYPTADAGVSJ--DYVGKADP 496
 OY 662 FDLSTRLEHIEHSIIONGMNR-----TYSRYRLDKLTKQAPPEWQDLP--VDVF 713

DB 497 KKAISTIKOYTTTA-----GAGIRMSVPVTEYDRNNEGLVAEHLTVNKNRAPHYADFI 552
 OY 714 N-GK-PSOALLAGVAVHKTV-----ADNLVMPMRGYRQYSLVSSGLVSDANAI 764
 DB 553 KKYKTDGCTDGSFPGWMLXKGTVGCGRKNTDSALMPTRGY-----LGVNAEI 599
 OY 765 ARAISGVYSPGDNAVYSNRAHQMTG-----GIQAGYIWSDNFNHVPYR 808
 DB 600 ALPGSKLOY-----YSATHNOTWFEPLSKTFTMLMGEGVIAGGY--GRTKEIPF 648
 OY 809 LRFPGGDSIRGVAHNDSLSPISDKYLTGGQVLAAG--TAENYEPM-----KD 856
 DB 649 ENFYGGGLGYSRGYESGTLG--KYDEXGEKISYGKNKANYSALEPPMPGAKDART 705
 OY 857 LRLAVFDGINAYD-----KGFNDTKIAGGVYRMAS 889
 DB 706 VRLSLFADAGSVMDGKTYDDNSSATGGRONITGAGNTHKSTFTNELRYSAGAVTWLS 765
 OY 890 PVGOVRVDAVATYKE--EGNPRIKHPFICTPF 919
 DB 766 PLGPMKFSYAVPLKKRPEDEIQRFOGLGTF 797

RESULT 12

08VQ24 ID 08VQ24 PRELIMINARY; PRT; 798 AA.

AC 08VQ24;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Omp85.
 GN Omp85.
 OS Bartonella henselae (Rochalimaea henselae).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bartonellaceae; Bartonella.
 OX NCBI_TaxID=38323;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Zimmermann R., Augustin K., Schaal K., Sander A.;
 RT "Cloning, nucleotide sequencing, and expression of a hemin-binding
 protein of Bartonella henselae".
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF461795; AAL66374.1;
 DR InterPro: IPR00184; Bac_surfAg_D15.
 DR Pfam: PF01103; Bac_surface_Ag; 1.
 DR SEQUENCE 798 AA; 88985 MW; 96A3E9C9066B7084 CRC64;

Query Match 5.4%; Score 254; DB 2; Length 798;
 Best Local Similarity 21.9%; Pred. No. 1.7e-06;
 Matches 203; Conservative 136; Mismatches 352; Indels 236; Gaps 52;

OY 96 SRIGQSPPLGDMGVIEETPL-----SLEELFAQESTEMGINPNDYIPEYGEQPNSEY 151
 DB 5 SRFLLAASVLYGMVIAPTTAFTAVSIAMVEEYQAVNSIEVHKNFY--SAGVVDNDTI 63
 OY 152 VVPPTLEP-EKPGILKRLYA--RLFND-GVNVY-PRU-----KAKFYOSSGSETS 197
 DB 64 KYGKSFSSGDDVFAVKRLPALGLFYDVKINGDILVLYKEYEVVNOYLVFGNKSLEDP 123
 OY 198 AIGS--SHOKTEPYANITAALEDITOESAMDINGSIPRLROTALVAARAVGYDILSII 255
 DB 124 DLKREISLKPNEPENSAR-----LSADVNIARE--AYKTVGRNDIAVITQ 166
 OY 256 RNSIGSEVDY-IIHDLGEVYIDYRAVEVRGEGA--DQKAFITYADEVP-----LLIGVF 307
 DB 167 TINLCKGRVNVFNVBSRRKISNITFKGNHAGSSRLRDYISTKPSGILSLLRGDY 226
 OY 308 HHGYYETKKNLEENASAEHGFDCGMLDRSDVILPDTADVSLIYDTGYRDEVVFF 367
 DB 227 SEERLADEALRFRFYNYRGYADFRVYSSKAVFPDEARNAEYIDVLYDBGVYKIKGIV-- 283
 OY 368 TIDKTNQLTTPDKLPRKRELLBOLLTVNNGEAY--NLQAVRALSNDLIATRYFNWVN 424


```

OY 699 QAPPEYMODLPVDFVNGKPSOEALLAGAVAHKIVADNLVNMGRYKORYSLEVGSSGLVS 758
DB 391 EGA-----ARELSKALVPGFAWTRRDVDDLPFRGRNITLTGCAVAKGLLT 437
OY 759 DAMMATIRAGISGYSSGDMANYSNRAH-----QMTGGIOAGIYMSDNHNHVPRLRFAGG 815
DB 438 DOSFVNSYSYIRQYVPVQGDIFVARAEALGAVLTGAADG-----VPATLRFRTGG 488
OY 816 DOSIRGYAHDLSL-----SPISDKGYLTGGCVLAVGTAENYEFPMKDLRLAVFGDIGNAY 869
DB 489 TQIRGYDFQSIGNDVNGSLTPKRLVLTGG-----VEYQRMFLPMGGAAVFMDGTAT 541
OY 870 DKGFYNDTKIGAGYGVWRASPVGVQVAVATGYKE 904
DB 542 DNMWTRRMFNGVGVYWRKSPVPIQLDLAVGIQO 576

RESULT 16
O8UC83
ID 08UC83 PRELIMINARY; PRT; 641 AA.
AC 08UC83;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DN Hypothetical protein Atu2615.
GN Atu2615 OR AGR_C.4742.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphlammachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Hounmel K., Gordon M., Vaudin M., Iarchouk O., Epp A., Liu F.,
RA Mollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cleo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AE009208; AAL43596.1; ALT_INT.
DR EMBL; AE008175; AAK8337.1; -.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 641 AA; 68281 MW; ECOA11DA3BDBF61E CRC64;

Query Match 5.0%; Score 236.5; DB 16; Length 641;
Best Local Similarity 20.3%; Pred. No. 1.3e-05;
Matches 162; Conservative 107; Mismatches 246; Indels 283; Gaps 31;
OY 200 GSSHOKTEPYANIKALEITQESANDLNGSI-PRLOTALVAVARAVGYDIDLSIRNS 258
DB 49 GSEPEVEYINPVKVA---VTIDAA-DADKSLKSSLENSLLIADKDRASGDLGLITRA 104
OY 259 ICEVDVYIHDLGEPV-YIDYRAVEVRGEGADDAKFTTVADE---VPLLT---GDVFHHG 310

```

```

DB 105 RDDDRLLIAALYENARAGGIYVNTVAGKNVDLPNPFVDFSTVPVMTVTPGPKFTIG 164
OY 311 KYE-----TKKNLJNNSAENHGYFDGRMLDRSVVILPD----- 344
DB 165 NVRLGQVYGNRLNDEYGLIAGDGAGSLAIRAGKKLIDDLKAEGRPLAKLTRREAVANHA 224
OY 345 -NTADVSLYDGTQYRFDEVEVFFIDPKTNQLTTPDPKLPVKKRELEQLLTVMNGEAYN 403
DB 225 TMTVDITMAEGGVAVPLGANTV-----TGEKIVDGD-----FIRYSRNLNGGEYS 271
OY 404 LQAVRALSNDLIATRYNNMVTETVPEREQIQNDQVSFEQSSSRPEPAQVDESTLEPV 463
DB 272 PEKLRKAADRLRLQGVF-----SSLFKEAGTLARDGTIPL 307
OY 464 IETVELDGIIMDSPIEFSSASN.IODKLNLYAAKARHLYMPDRRLAIHHDGV---- 519
DB 308 -TIEVSEG-----KHRYFGVAGY-----STTEGIGLOG 335
OY 520 ---NRSILG-----RISDAVASAVARAILPDESENEVIDLPE-----RTA 555
DB 336 YMGHRNLFQGAESLRIGSVSRIAEA-----SVEGDYSAGITFTYPCMPNPTTKRTS 390
OY 556 LANRKTADYQSKKVPDYFVVASDKPRDQIGLWGSDTGTRLVTKFEHNLINRDGYOA 615
DB 391 LIAKTENPDYRAKTL-----TGT---AGPAYELNDTIDTAA 424
OY 616 GAELRL--SEDKKVKLYATKPLSHPL-----NDQLRATLGYQOEYFGHSTNGFDLST 666
DB 425 GLEVQWADTDEDAFGKNEYLT--TSIPLEFYADTDDKLNPTGFRAS----- 469
OY 667 RTLEHEISRSIIQNGMNRFRYSRLYRDLKLTQAPPEYMODLPVDFVNGKPSOEALLAGV 726
DB 470 -----IAAKSYEAL----- 479
OY 727 AVHKTADVNLVPMRGYKORYSLEVGSSGLVSDANMAIARAGISGYSSGDMANYSNRAH 786
DB 480 -NOTFFSSSEGSITGYK-----GLGAEDRLIM-----AG 507
OY 787 QMTGGIOAGIYMSDNHNHVPRLRFAGGQOSIRGYAHDLSPISDKGYLTGGCVLAVGT 846
DB 508 KLSGCVLVG---GSDLDIDITTRRFFAGGGSGVGSYQETSIPYANAGDAGTGRSVYGS 564
OY 847 AEYNYEFPMKDLRLAVFGDIGNAVDKGFT---NDPKIAGYGVWRASPVGVQVAVATGYK 903
DB 565 VEARIKVTDTITGLVPFDACVSD-GVTPDFSDIRAGAGIGLRATPFGPLRLDVAMPLE 623
OY 904 --EGNPIKLHFFITGTF 919
DB 624 KYDGGNNFGIYAGIGOSF 641

RESULT 17
O8UFL6
ID 08UFL6 PRELIMINARY; PRT; 774 AA.
AC 08UFL6;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DN Group 1 outer membrane protein.
GN OMP1 OR ATU1381 OR AGR_C.2554.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,

```

Raymond C., Rouse G., Saenphimmach C., Wu Z., Romero P., Gordon D.,
Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
Neeter E.W.,
RT "The genome of the natural genetic engineer *Agrobacterium tumefaciens*
C58."

RT Science 294:2317-2323(2001).

RL [21]

RN SEQUENCE FROM N.A.

RX MEDLINE-21608551; PubMed-11743194;

RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
Qurollo B., Goldman B.S., Gao Y., Askenazi M., Halling C., Mullin L.,
Houmel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
Mollam C., Allinger M., Doughly D., Scott C., Lappas C., Markelz B.,
Planagan C., Crowell C., Guron J., Lomo C., Sear C., Scrub C.,
Cielo C., Slater S.,
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58."

RL Science 294:2323-2328(2001).

DR EMBL; AE009099; AAL42387.1; -
EMBL; AE008064; AAK87173.1; -
KW Complete proteome.

SO SEQUENCE 774 AA; 84546 MW; 3C88B25F4EB537DF CRC64;

Query Match 4.9%; Score 232.5; DB 16; Length 774;

Best Local Similarity 21.0%; Pred. No. 3e-05;

Matches 178; Conservative 109; Mismatches 310; Indels 251; Gaps 39;

OY 165 IKRLRYARLF-----NDGVNKV---PRLKAFYSSGSGGSAIGSS 202
DB 77 VKRLYATGVSNVSMRVSGSTLVVTVNENQLVNQNQVNGNRKIKDKDLAGIVQTPMGPF 136
OY 203 HQK--TEPVANIKAALEDITQESAMDLNGSIPRLROTALVAARVGYVDLSII-- 255
DB 137 NQIVAVADIARIEVAISIR-----SDVEITTOIV-----SVGGGRVNIAPVINEGE 184
OY 256 RNSIGEDVVIHDLGEVYIDYR-AVEVRGEGADKAFITVADEVPLIDVFNHGKYEY 314
DB 185 RTKIGRIDFT---GNNSYSDGRILAIVNTKKSNMISFLRK-----DVINEDKLRA 232
OY 315 KKLILNANASAEHGYFDGRMLDRSDVILPP--NTADVSLIYDMGTQYREDEVVFTIDPK 372
DB 233 DEBALNQFYVNRGYADFRVY--SSDAVYLDSESKNEYTISIVDEGKAYDFGNVAVESTVP- 289
OY 373 TNDLTPDPKLPYKRELEQLTVNNGEAVNLQAVRALSNDLATRYFNVNVEIYFPER 432
DB 290 -----GVDSLELQGLVETTRGASTSAKEVQO-SNEAISKR----- 323
OY 433 EQLONDVSFEQSSSRTEPAQVDESTLEPIETVELTDGILMDISPIEFSASNLIDOKL 492
DB 324 -----VAGEGYFPAVTPPRGDRDMSGMTIGYTVIVDGERAYVERIELRGTRTRIDYV 376
OY 493 NLYAAKARILYDMPDDRYLAINHDDGVNRSI-----IGRISDAVASAVARATILPDE 542
DB 377 -----IRREFD-----ISEGDAFNQTIITAAKRLLEALGYFSKYNISVAGSAPD- 421
OY 543 SENEVIDLPERALANRKTTPADVOYQSKKVPYVFAVSDKPRGQIGLGMGSDTGRFLYVK 602
DB 422 -----RVYIVDVEDOSTGSFGIGAGYSQNDGVILLEAS 454
OY 603 FEH-NLINEDGY--QAGAELRLSEDKKGVKLYATKPLSHPLNDLORATLIGV----- 651
DB 455 VBEKNFLGQYIRVAAGG--EDDARTYLSITFEPI-----FLGTRLAAGFDL 501
OY 652 -----QEVGHSTNGEDLSTRTLEHETSRISIIQ-----GGMNRTY 687
DB 502 FKNQSKSDSYVNYDEQGFAL-----RYTAPITENLSTTFKYTYKQIYBECKGDMQNA 554
OY 688 SLRYRLDLKLTQAPPEYMODLPVDFVNGKPSQEBALLAGVAAKTYADNLVNPMSRYRORY 747
DB 555 NL-----AEPYOAL-----IRGEDWTQSTLSMTLVNNTLDBRM--PREGMOAL 597

OY 748 SLEVGSSGLVSDANNAIARAGISGYVSFC--NATGNSRAHDMGTGICAGTWSNENHV 805
DB 598 TNEF--AGLGGSEYKTYAKARYTYTISDEYDVS-----LFG--QAGHVMPTGMDL 648
OY 806 PYRLRFPAAGDQSIIRGYAHDLSPI--SDKGYLTGGQVLAVGTAEYN-----YEFKMD 856
DB 649 VED-QFKFGGRQ-VRGFANDIGIPRISDS---IGTITFYAASAEVTPMPGVEDE--G 701
OY 857 LRLAVFGDIGAAYDKGFT-----NDTKIGAGVGRWASPVGQVRVATGKVEEGNP 908
DB 702 LRLAGFVADAGTMYGNKYSTQTVKDDNSIRASAGIGVMAASPFGPIRVDAVPIAKEDYD 761
OY 909 IKLHFFIG 916
DB 762 EQQRFREG 769

RESULT 18

O93PM2

ID O93PM2 PRELIMINARY; PRT; 793 AA.

AC O93PM2;

DT 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Integral outer membrane protein.

GN D15.

OS Haemophilus ducreyi.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

OX NCBI_Taxid=730;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-21295095; PubMed-11401984;

RT Thomas K.L., Leduc I., Olsen B., Thomas C.E., Cameron D.W., Elkins C.,

RT "Cloning, Overexpression, Purification, and Immunobiology of an 85-

RT KiloDalton Outer Membrane Protein from *Haemophilus ducreyi*."

RL Infect. Immun. 69:4438-4446(2001).

DR EMBL; AF329831; AAK70345.1;

DR InterPro; IPR000184; Bac_surfAg_D15.

DR Pfam; PF01103; Bac_surface_Ag; 1.

SO SEQUENCE 793 AA; 88661 MW; E4E9DB62A8F903F CRC64;

Query Match 4.9%; Score 229.5; DB 2; Length 793;

Best Local Similarity 19.3%; Pred. No. 4.7e-05;

Matches 173; Conservative 134; Mismatches 309; Indels 279; Gaps 44;

OY 175 DGVNK-----VRLKAFYSSGSGGSAIGSSHQKTEPVANIKAALEDIT----- 220
DB 28 DGQVTEGNAIILASLVYKVGQTTDGDVTRIVKQLFSQNRFFENVASARSGQTLVIVAE 87
OY 221 ---QESAMDLNGSIPR-----LROTALVAA-----RAVGYDIDL 252
DB 88 PLINLITLIGNNALIPKNALEONKANLIVAGEYDKAKLEAFQALVDHYHTMGRYQADI 147
OY 253 SIIR--NSIGEDVYIH-DLGEVYIDYRAVEYRGEGADKAFITVADEVPLIIG----D 305
DB 148 QTTTPPNNGSINVELINTEGETAYVK---KINFGNNAFSDDELKELEIKPAPMWN 203
OY 306 VFHNGKYETKK-----NLLENASAEHGYFDGRWRDSDVYILPDRNADVSLY--DTGQ 358
DB 204 IFESSKFOQDEYNKDEILRDYMDHG--AKFTLKDVTVKFNENKTEVDLYKKINEGSO 261
OY 359 YRFDEVVFTIDPKTNQTLTDPDKLPYKRELEQLTVNNGEAVNLQAVRALSNDLIATR 418
DB 262 YNISEK-----RIIGDTQKLDNE--LNQLLT-----HFKAQQ 291
OY 419 YFNMYNTELVPEREQIONDOVSFEQSSSRTEPAQVDESTLEPIETVELTDGILMDIS 478
DB 292 LFRKTELSTIEQIKQIIGIDR---GYGSAK-----VDLY 322
OY 479 PIEFSASNLIDOKMLVAAKARHLVDMDDRYLAINHDDGVNKRSTLGR-----ISD 529
DB 323 KFNEDHHTVO--INFIYDAGRRTY---VKKIREGNDVADSTLRREMQOEGAWLST 376

QY	530	ANSAVARALPDESENEVIDLERIALANKRTPADYVQSKVPLVYVASDKPRDGOI--	587
Db	377	SAVLSAKSLSEFTGFEYETVEVSMPT-VKNTDDOVDIYIKK-----ERNTGSI	425
QY	588	GLGMSDGTGRLVTFKEFNNL-INRGYOA-GAEI RL-----SEDKGKVLATKP-----	635
Db	426	GVCYSGSGS-----LSINAGITQDNFLGMSISGLGNSRNTDYNLSYTEPYFTKDG	479
QY	636	--LSHPL-----NDOLPALTGQOEVEFGHS-TNGFDLSTRTLEHSRSIIIONGGMRT	686
Db	480	VSLGINFIYEDYDNRASAKSAAYKRRTVYGASGTLGFPYDE-----NNS	522
QY	687	X--SLRFLYLDKTKTQAPPEETHQDLPVDFVNGKPSQEOALLAGVAVHKTVADNLVPMKGYR	744
Db	523	YLLGLGYHDHKLRYNREYRETB-----KVNMSKRP-----INPQSHY	561
QY	745	QR-----YSLKVG-----SSGLVSDAMMALARRGISG-----YSPQDNAG--	781
Db	562	DRIGQADDPDLSRGWYNNLNNGYEPPTAG---SSANISGKLTLPQSDNKKYQVGTNPSGYI	618
QY	782	-SNRAHMTGGIQAQYIWSDNF--NHVYRLRFPAGDQDQIRGAHDSLSP-----ISDK	834
Db	619	PLNSEHKWVIATKGGIAYTNSFGGKEVPEFYLYSAGGMSGLRGAAGSISPKAIYREDS	678
QY	835	YLTGGQVLAIVGTAEIYN-----EFMKDLRLANF-----GDIG	866
Db	679	FKAPSQDVIIGNANVMNLSLEILIPAPISDKYQHNVTSTVFVAATVWNTMKRQSKADPY	738
QY	867	NAVDEGFYNDRIKIGAGVGRNASPVGOVRVDAVGRK-EGNPI-KLHFFITGPP	919
Db	739	NLPDGDYKRRRARSAGIAMLQMSPIGLPSTSYAKPIKKYAGDEIEOPQFQVGSSTP	793
RESULT 19			
P73472		PRELIMINARY;	PRT; 861 AA.
ID	P73472		
AC	P73472:		
DT	01-FEB-1997 (TrEMBLrel. 02, Created)		
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Chloroplast import-associated channel IAP75.		
GN	IAP75 OR SLR1227.		
OS	Synechocystis sp. (strain PCC 6803).		
OC	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.		
OX	NCBI_TaxID=1148:		
RN	[1]		
RA	SEQUENCE FROM N.A.		
RP	MEDLINE=97061201; PubMed=8905231;		
RX	Medline=97061201; PubMed=8905231;		
RA	Kaneke T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,		
RA	Mayajima N., Hirosewa M., Sugijima N., Sasamoto S., Kimura T.,		
RA	Hosouchi T., Matsuno A., Muraiki A., Nakazaki N., Naruo K., Okumura S.,		
RA	Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,		
RA	Tabata S.;		
RT	*Sequence analysis of the genome of the unicellular cyanobacterium		
RT	Synechocystis sp. strain PCC6803. II. Sequence determination of the		
RT	entire genome and assignment of potential protein-coding regions."		
RL	DNA Res. 3:109-136(1996).		
DR	EMBL: D90906; BAA17512.1; "		
DR	InterPro: IPR000184; Bac.surfAg.D15.		
DR	Pfam: PF01103; Bac.surface.AG.1.		
DR	TIGRFams: TIGR00992; 3a0901s03IAP75; 1.		
KW	Complete proteome.		
QO	SEQUENCE 861 AA; IC50FED394DA5B44 CRC64;		

Query Match 4.8%; Score 225; DB 16; Length 861;
Best Local Similarity 19.7%; Pred. No. 0.0001;
Matches 189; Conservative 130; Mismatches 399; Indels 242; Gaps 38;

QY 93 SPIRIGESOSPILGDMVIEETPLS-LEELFAOESTFMGINPDNYIPYEGGEQNSFV 151
|| : | : | : | : ||| : : : : | ||
Db 11 SPYRLLLTSLVLGASPAQAQTAFSPLNMLADDNTELLAVPSSEL-NFGPPISESEA 69

QY	152	VPPPLTEPEKPELLIKRLARLFLNFCVNMKVPPLPKAKFFYSSOSSGSETSAIGSSHOKTEPYAN	211
Db	70	SNVSTFLOEAPBSVINALNGE---NGEISVTP-ETIEQIDGDLGGOTMELSA-----GN	118
QY	212	IKAALEDITQESAMD-----LNGSIPRLRQALVAARAVGYDIDLSTIRNSICEVDVI	266
Db	119	LDOGVDDILPRAPVDSALDAQANELPHONNAVOAPRAVOEGENLIAETKTDQOTDHV-	177
QY	267	HDLGER-----VYIDRAVEVREGGADKAFVTYVADEVPILIGDV-----FHNGK	311
Db	178	----EPOSPILMAQAAVEEVEVAEMEATEETTGVTETPEETPSFTDPADPTNTEGTPGR	233
QY	312	YETKNKLIENNSAEHGTYDGRWLRSDVYILPONTAD-----VSLYDTGTOYRF	366
Db	234	TOTLPSTPPAS-----PSTTPAPAREEBRVLVSEVLVYGTTPREL	274
QY	362	DEVEVFTIDPTNOLTTPDRLPKVRELLFQLLVNNGEAVNLQAVRALNSIDLATETFPN	421
Db	275	ELVYV-----MAIRTPGKTTTRTQOLEDV-----NAITATYFS	309
QY	422	MVNEIVEPPEBEOIONDVSFEQSSSS-----RTEPAQ-----VDESTLE-	461
Db	310	NVR---VAPSDTPL-GVAVTEPEVOQANPVFTGLNIRTPETAEGKERILLPQEVDETGEQ	365
QY	462	--PVTEIYELTDGLIMDISPIEFSASNLIDOKLNLVAAKARHLKYMPDDRYLAINHDOV	519
Db	366	YKGIIMLHELOEG---IKTINENYSNOGYDLAOVGSP---QVADGQVTLVIAEGI	416
QY	520	NRSILGRISDAVSAVARAILPDESENEVIDLPETALTANRK---TPADVYQSKVY	571
Db	417	VENIOVRFED-----SEDEVOGRTGDFITITREMRLLKPDVFNRNNAQTDLQ	463
QY	572	-----PLYVFVASD--KPRDQOIGLQWG--SDPTGRLVTKFEHNL	607
Db	464	KVYSLGLEFDYRLSPNFGSDPTEYIVAVDVEGNTGSLAAGGSSSSGLFTIISYGERN	523
QY	608	INRDGYQGAELRLSEDKKGYKLYATYRPL--SHPLDQLRATIGYQOEVEGHSITNGFDLS	665
Db	524	LGGNMOTIGVAAQVGOHELLFDVYSFTDPMWIGGDPEFRTSYTANL-FRRRTTSLVEFGDASS	582
QY	666	TRLLEHISRSIIQNG-----GMNFTYSLRRRLDKLKQAPPEPMQDL	708
Db	583	IRTNFGPDSPPRVVTRTGGLTFEPRADIADVAPRDPWRLSAGFGYQVNNAIENAAALSPFSA	642
QY	709	PVDFVNGKP-----SQEALLAGVAHKTIVADRLVNPMPGYRQORYSLEVGGSLGVSDAN	761
Db	643	PLNFNFSQPLSFSDYGYVDELFTLSFGASQDNRRNNALOPTGSLVRGAE--QITTPGTGN	700
QY	762	MAIRAGISGVYSF-----GDNAVYSNRAHOMTGGIGQAGYIWSDNFNHVPYRLRF	811
Db	701	IMMTR--LRGYSYIIPVNNMLDLTGLFGLVBSSTQPOVAFVNAQGVTLGD---LPPYEAF	754
QY	812	FAGGDQSIIRGCAHNSLSPIDSKGLTNGGCVLAVGTAEYNAENEFMKDLRLAVFGIGN----	867
Db	755	ILGSGNSVVRG-----QEEBELNGRSFFQATAEYRPIILAAVGALFVYVGSGNLGS	805
QY	868	-----AYDKGFTNDTKIGAGGVWMAPSGOVARDVATGVKKEGNGPIKLHFFIGTGR	919
Db	806	QGAVPGEPAIYKRLG--SGVGYLGLAVRTQSPVPIRIDL--GTTGEGES-RINFGIGEXF	861
RESULT 20			
Q8YHHO	Q8YHHO		
AC	O8YHHO	PRELIMINARY;	PRT: 781 AA.
DT	01-MAR-2002 (Tremblrel. 20, Created)		
DT	01-MAR-2002 (Tremblrel. 20, last sequence update)		
DT	01-JUN-2002 (Tremblrel. 21, last annotation update)		
DE	Outer membrane protein.		
GN	BMEI0830		
OS	Brucella melitensis.		
CC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;		
CC	Brucellaceae; Brucella.		

OX NCBI_TaxID=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / Biotype 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA Delvecchio V.G., Kaprielian V., Redkar R.J., Patra G., Muijer C., Los T.,
 Ivanova N., Anderson N., Bhattacharya A., Lykidis A., Reznik G.,
 Jablonksi L., Larsen I., D'Souza M., Bernal A., Mazur M., Goldsman E.,
 Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Lelesson J.-J.,
 RA Haselkorn R., Kyriplides N., Overbeek R.,
 RT "The genome sequence of the facultative intracellular pathogen
 Brucella melitensis";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL: AE009524; AAL52011.1;
 DR InterPro: IPR000184; Bac_surfAg_D15.
 DR Pfam: PF01103; Bac_surface_Ag1.
 KW Complete proteome.
 SO SEQUENCE 781 AA; 85918 MW; E3C7385CCB11DB3A CRC64;

Query Match 4.7%; Score 223.5; DB 16; Length 781;
 Best Local Similarity 20.9%; Pred. No. 0.0001;
 Matches 180; Conservative 116; Mismatches 302; Indels 263; Gaps 45;

OY 165 IKRLIA-----RLFDGVNKKVPRLLAK-----FYOSSGSETSAIGSSHQKTEPYAN 211
 DB 78 VKRLFAAGLPSDVIRHSGSTLIQVQKERSVNNVLFQGNKKIKOPDLARAQV-LKPRAP 136
 OY 212 IKALEDITQESAMDLNGSIPRLRQTLVAARAGYDIDLSIRNSIGEDVYIHDLGE 271
 DB 137 FDMATMAOKAEIKAIVSHIG--RSDATVAKRV-----DL-----GQGRNV--VYEINE 183
 OY 272 PVYIDRAVEVRREGADAKFT--TVADEVP-----LLIGDVFHHGKYEKKMLIE 320
 DB 184 GSRTKIANIEFVG-----NQAFSGRLRDVISTKRSNPLMLTRNDVYDEGRLOADETLR 239
 OY 321 NASAEHYFPGCRMLDRSDVYILPDNTADVSLITDGTQYRPEV-VFTIDPKTNQLTTD 379
 DB 240 RFTYNRGYAFRLSSNAVLDPSTNEYTITITVDEGPRYFGDVSVEJVD----- 290
 OY 380 PDKLPVARELEQLLVNMGAVNLQALNSDLITRFNMVNTETIVEPEREQIONDQ 439
 DB 291 ----GVDTQALDLVLRTRTKPKISAKIE-----DSVLSTESVAGSG 329
 OY 440 VSEQSSSTEPAPQVDESTLEPIVETVELTD-GILMDISPEFSANLQDKLNLVAAK 498
 DB 330 YAF-----AKVEP-RGDGRNFENHTISVYVSVDGPRAYIORIRIGDKRDYV----- 377
 OY 499 ARHLVMPDRVLAINHDCVNRSLIGR-----ISDAVASAVARAILPDESENEVI---D 549
 DB 378 IRREFD-----LNEGDAFNQVAVQRAKRRLLEVLDFFQTVNISTAPGSEPDQVILVVD 429
 OY 550 LPERTALANRKPADVYQSKVPLVYFVASDKRQDQIGIGMSDQG-----RLVLT 601
 DB 430 VVEKST-----GEFSISGGYTTGGESPGAQVEAALIT 460
 OY 602 KREHNLINRGY--QAGAELRLSEDKGKVLATKPL-----SHPLND- 642
 DB 461 --ERNFLRGQYIRISAGAC---QDDMRNVGLSFTEPYGLRSLAGFDVFRSYVND 515
 OY 643 ----QLRATLGIQOEYFVGSTNG-----FDLSTRLEHISRSIT---QNGSMNR- 685
 DB 516 YDEVTGCGTIRGLPTTDFNSAGIAYSLVQEKYDILFRGDAENVYAPALLEAENSFWLS 575
 OY 686 --TYSLAR-RLDKLTQAPETMODLPVDFVNGKPSQE-ALLAGVA--VHKIVADNLVMP 739
 DB 576 SVSYSLTITSSIDIKN-----PHDGLGKFIQEPAGLGDAKYKTKFKGK--- 621
 OY 740 MGRYQRYSLVSGSLVSDANMAIARAGISGVYSGFDAVYGSNRAHQMTGGIOAGYIMS 799
 DB 622 ---YQTLISQEADIVGLG-----VGAGYIHFGDD-----GVRT----- 653
 OY 800 DNFNHPVRLRFPAGGDOSIRGYAHDSLSPISD-----KGYLTGGQVLAVGTAEVNEEF-- 853

DB 654 -----FDLFKNSDILRGFKFNGIGIPYODAKNKGRIYMGTTYSGAIEVOPMPV 704
 OY 854 ---MKDLRAVFGDIGNAVYDKGFTNDT-----KIGAGYGV--RMASPYQGVENVAT 900
 DB 705 LPESLGVGAFFADAAATLYG-----NDTPDISGDDKKLRASVGVSLMMSPFQPLAFDVAF 760
 OY 901 GY-KEEGNPIK-LHFFIGTPP 919
 DB 761 PVAKADTKDVQNFNGVSTKF 781

RESULT 21

ID Q98MC3 PRELIMINARY; PRT; 794 AA.
 AC Q98MC3;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE Outer membrane protein.
 GN ML0636.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=MAF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Matanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsunoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti";
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP002995; BAB48190.1;
 DR InterPro: IPR000184; Bac_surfAg_D15.
 DR Pfam: PF01103; Bac_surface_Ag1.
 KW Complete proteome.
 SO SEQUENCE 794 AA; 85741 MW; BABE4A1BC8FEEDC6 CRC64;

Query Match 4.7%; Score 223.5; DB 16; Length 794;
 Best Local Similarity 20.4%; Pred. No. 0.00011;
 Matches 177; Conservative 130; Mismatches 321; Indels 241; Gaps 44;

OY 156 TLPEPK-----PGLIKRLYAR-LFNDGVNKKVPRLLAKFYOSSGSETSAIGSSHQKTE 207
 DB 62 TIRPKAFSSSDVDAVAKALFGTGLFSD-----VQINQVGSITLVVVAEKVV 109
 OY 208 PYANIK--ALEDITQESAMDLNGSIPR--LRQTA-----VAVARAGYDIDLSII 255
 DB 110 NQVLFQGNKKLKNALAAVQLK--PRGTFSGQATLDSDEAVKAAVRRIGRDAGVTTQ 166
 OY 256 RNSIGE--VDVLIHDLGEPYIDYRAVEVRGEGA-----DKAFTTVADVEPLLT-GDV 306
 DB 167 VVELGDNRVNVFVH-TTEGDRQIQAIAINFEVGNAYSRRSLSDVINTRKSSWVSFIIRDYV 225
 OY 307 FHHGKETKKNLLENASAEHGYDFGRMLDRSDVYILPDNTADVSLYDGTQYRPEVVF 366
 DB 226 YDEDKLRADQELLRRYYNHGADPQVYSAVGEIDNATKYYTITVQGERINFQDISV 285
 OY 367 FTIDPKTNQLTTDPDKLPVARELEQLLVNMGAVNLQAVR---ALSNDLIATRY-FN 421
 DB 286 ESTIRE-----VDSKSLSESVETHHGQDVYNAKDVEDSITALLEKVAAGSYATA 333
 OY 422 MNTETIVEPEREQIONDQVSEQSSSTEPAPQVDESTLEPIVETVELTGDILMDISPIE 481
 DB 334 QVT-----PRGRD-----NFENHTIS-----VVYITVD--QGTKAVYIERIE 366
 OY 482 FSASNLIDQKLNLAAKARHLVMDPDRVLAINHDCVNRSLTGLKISDAVSAVAARILPD 541
 DB 367 IRGNDRTRDYV-----IRREFD-----VSEGDANQVLIOR-----AKRLED 404

```

0Y 542 ESENVIDILPERTALANKRTPADYQSKKVPYLFVNSADKRRDQOIGLQMSDTCGLVLT 60Y
Db 405 LMYFKVEI-----STVPS- -QPQVVLVVDYVEKSTGEFSVAGY- -STGDSAGP 453
0Y 602 KEHNILNRDGYQAGAEILRLS-----EDKKGKVLATKPLSHPLDLQJLATLGVOEFGH 657
Db 454 SVESGISTERNFNLGRQPIFKLSAGGKKNRDYSVSPTE- -YFLGRILAA----- 500
0Y 658 SINGNDLSTRLEH-----ISRSITQNGKNRITYSLKTRDLKLTQA 700
Db 501 ---GDIIKSTREYNANNYDSDPTGATVREFGLPTNSITTOATVNIISOE-KKYVD- -NSC 553
0Y 701 PPEYQDILPYDFVFNKGPSQOELAGVA-----VHKIVADNLVPMRDEORY 747
Db 554 GPTDPNPNPGTCN---ISQALDIQSIASPMIKSSVSLGLVYNTI- -DMMKNPHGII- - 600
0Y 748 SLEVGSSGLVSDANNAIRACISGVYSF-----GDNVAGSNRAHOM-----TGGIOAGYI 797
Db 607 -----ANTTVAVGAGVGCAKPVKITTGRSIIQYTLSEOLDVLGLISG- -AGHV 652
0Y 798 ---WSDN-----FNHVPYLRFPAGGDSIIRYAHDSLSPID--KYLTGGQVLAVTAE 848
Db 653 EGYGDGDLRIFDH-----FQSTDRMRIRKFPAGGIGIPVASGTSGHLLGGTYVFNASAE 705
0Y 849 YNYEF-----MKDLRAVECDIGNAYDGFNFNDT-----KIGAGGVEMASAPVG 892
Db 706 AQFPLVPVIPSFGUGAVFADAAATLYGKIKIADQTLVDDQSTGCMKLRASVSGVGLMMASDPFG 765
0Y 893 QVRVAVATGVKKEEGNP--IKLHFFIGTFF 919
Db 766 PRIDYAIPIVREASDDVOEENFGISTEFF 794

```

QJ	SEQUENCE	615 AA;	67775 MW;	B9A64607638E1CCB	CRC64;
QJ	Complete proteome.				
KW	EMBL: AL162752; CAB83602.1; .				
DR	NCBI_TaxID=658699;				
OX	NCBI_TaxID=658699;				
OC	Bacteria: Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.				
GN	Neisseria meningitidis (serogroup A).				
DE	Putative outer membrane protein.				
DT	01-DEC-2001 (Tremblrel. 19, last annotation update)				
DT	01-OCT-2000 (Tremblrel. 15, last sequence update)				
AC	QJ9JW5;				
ID	QJ9JW5	PRELIMINARY;	PRT;	615 AA.	
RESULT 22					

Query Match	4.78	Score 220	DB 16	Length 615
Best Local Similarly	22.38	Pred. No. 0.00011		
Matches 131, Conservative	72	Mismatches 236	Indels 148	Gaps 19

```

0Y 462 PVIEFVELTQGLIANDISPIESASNLQDKNLVAAARHLXOMPDPRVALIHHDSVNR 521
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 48 PDTESVKLKPAPRIPRIDTQSEIKDWEHLPLITQOEEVLDEKQTGFLEAAPDNVKT 107
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
0Y 522 SI--LGRSDAVSA-----VAKALLPDSESENEVTDLPERRALAN 558
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

D	b	108	MLRSKGYSSKVSILTEKXGAVTVHTTGPRTKIANVGAIIIGDLSGNI,LAETYYRANLEN	167
O	y	559	RKTRA-----DYQSK-----	562
D	b	168	WOQPPGSDDEDDSSWENKTSYLGAVTRKAYPLAKIGNTRAUVNEDTATADLNVVYSSGRP	227
O	y	583	RDGQIGLGMCSGDTQTR-----LVTKFEPHNLINRDGYOAGAEI-	619
D	b	228	-----IAPDEFITGTQRPQPEQIYSGLARFPQRPQPYDDLLDLPQAL-EDONGHSGASVQ	282
O	y	620	-----RLSDDKKGVKLATKPLSHPLNDQLRAT-----LG-----YOEVFPGHSTNG	661
D	b	283	ADFPRLQGDPRPVVSVSYTEVRKHLEGTIRLDESEYGGKIAADYVNLFNKGIGSVYMD	342
O	y	662	FDLSTRTEHIESRSIIIONGMM-----NRTYSRLRYBLDKLTQOAPETQODLPVDFVNGKPS	718
D	b	343	MDKVTETIAGISQPRNTRKGYWTVSNVSYNRSTQONEKRAFSGSIVYVDRAGIDRRLG	402
O	y	719	QEALLAGVAV-----HKT-----VADNLVPMRGYRORYSLEVGSSGLVSD	759
D	b	403	AEEFLAEGKKRIGSDIDIGNSHATMLTASMKRQOLLNLYLHPENGHYLDGKIGTITLGAFLS	462
O	y	760	ANM--ALARACISGVSPGDMAYSNNAHOMTGI-----QAGV-IWSDNHNHYPLRLREF	812
D	b	463	TALTRTSARRA-----YFTPENNKKLGFTPIIRGAGAGTVADARN--VPSGIMFR	510
O	y	813	AGGDOSIISGVNHDLSPIISDKGYLTGGQVLAAGVAEYNEBFMKDLRLAVGDIJANVDKG	872
D	b	511	SGGASVAVGVEYELDISGLAGRGVSLPERALLVYSLEQVLPFTRLISGAVPHMDGDAANF	570
O	y	873	FTNDTKIGAGYGVWVWASPGVGOVRYDVAIVTKYKEBGNPKLHFFIGTTP	919
D	b	571	KRMKLKHGSGGLGVWVFSPLAFEFSDIAYGHSDK--KLRMHSISLGTFR	615

RESULT 23			
098K30			
ID	098K30	PRELIMINARY;	PRT; 617 AA.
AC	098K30;		
DT	01-OCT-2001 (TREMBLrel. 18, Created)		
DT	01-OCT-2001 (TREMBLrel. 18, last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, last annotation update)		
DE	Hypothetical protein mll1662.		
GN	MLL1662.		
OS	Rhizobium loti (Mesorhizobium loti).		
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;		
OC	Phyllobacteriaceae; Mesorhizobium.		
OX	NCBI_TaxID=381;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-MAF303099;		
RC	MEDLINE=21082930; Pubmed=11214968;		
RA	Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,		
RA	Watanabe A., Igesawa K., Ishikawa A., Kawashima K., Kimura T.,		
RA	Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,		
RA	Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,		
RA	Takeuchi C., Yamada M., Tabata S.;		
RT	"Complete genome structure of the nitrogen-fixing symbiotic bacterium		
RT	Mesorhizobium loti.";		
RL	DNA Res. 7:331-338(2000).		
DR	EMBL; AP002998; BAB48984.1; -;		
DR	InterPro; IPR000184; Bac_surfA_D15.		
DR	InterPro; IPR003975; Shal_channel.		
DR	Pfam; PF01103; Bac_surface_Ag; 1.		
DR	PRINTS; PRO1497; SHALCHANNEL.		
DR	Hypothetical protein; Complete proteome.		
SO	SEQUENCE 617 AA; 65759 MW; 83BF8C176201A44 CRC64;		

Query Match	4.7%	Score 220;	DB 16;	Length 617;
Best Local Similarity	22.2%;	Pred. No. 0.00011;		
Matches 150; Conservative	95;	Mismatches 246;	Indels 186;	Gaps 30;

QY 313 ETKKNI--ENASAEHGYFDGRWLD RSV DVI, P DNTADVSLIYDTGTQYRFEDEVFFTI 369


```

Db 49 DADKDLVRLENASALKGED--RPVSGSLGLAKARSDREQLVAALFADAREGVTTIT 107
Oy 370 DPK-TNOLTTD-----PKLPVKRELBELLVNNGEALNLOAVRLSD---LATRF 420
Db 108 DCKPLDELPPAEPFGOPITV-----TINIAVGPFTLTIR-LEDDAAGLMSADVG 159
Oy 421 MNVNT-----IVPPERQ-----IONQVSFEQSSSRTEPAQVDESTLEP 462
Db 160 LISGDACSGAVALKAELIVRTLKEGRPLAKVTDRRIADATSTLDVTLTV---AAGP 216
Oy 463 V-----IEVELTDGILMDI-----SPIEFSASNLIDOKLNLVAAKARHLYDM 505
Db 217 VAGGATTVTEGKEVDKDFTEYMTGLKRGKQVSPQEISDAR----- 257
Oy 506 PDDRVLAINHDDGVNRSILGRISDAVSAVAAILPDESENEVIDLPERTALANKRTPADV 565
Db 258 --DRLLALE---VFNSTFK EADKLD A----- 279
Oy 566 YQSKKVPPLVVFASDPKPRDGOIGLGMGSDTGTRLVTKFEHNLINRDGYQAGAEIRLSEDK 625
Db 280 --DGNIPIGVQVSEKRPYFGLGTFSTBEGLLEGYMGH---RNLFGHAKEKLIDGAI 333
Oy 626 KGVLYLTKPLSHPLNDOLRATLTGYOE---VFHSTNGFDLSRTLEHEISRSITQNG 682
Db 334 SGIG-----SNNLSD-LANNAGIMEKPGVIGPASKFFAGFKTVLEHPADYHFSYK 385
Oy 683 WNRYSRLRYRLDKLTKTOAPETMODLPVDVFN-----CKPSQEALLAGVAVHKTV--ADN 735
Db 386 ---STGLSTELDKQQTYS-----AEVALDYSRTIDAFKHTY--LIASVPLQYVYDNDNS 435
Oy 736 LVNPMRGYR-----QRYSLLEVSGSGLVSDANMAIARAGISGVYSEGDNAVSNRAHOMT 789
Db 436 RLNPTRGFRVLAYAEPSYDIMGSAFL-----KLKGBSAYQSIDTAKFPLAER-- 485
Oy 790 GGIGAGVIMWDFNHVYRLRFAGQDOSTRGVAHDSLPSIDKGYLTGGVLAVGTAEY 849
Db 486 --VAIGIVTGTGLQNVADRRFYSGGGSVGYVQVIGRPFYQPIGGLSF----- 536
Oy 850 NYEFMKDLRLAV-----FGDIGNAVYDKGFTN--DTKIGAGGVWMASSVGVQVRDVA 899
Db 537 -FETSVEMRIAVYDTIGIVFPVADGYSTKSVNFSVDKVGAGVLYVPFGPLRIDAA 595
Oy 900 TGVKEEGNPIKLHFFIG 916
Db 596 VPLNRDNDPHFGIYAG 612

RESULT 24
O92LN9
ID 092LN9 PRELIMINARY; PRT; 618 AA.
AC 092LN9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein R02996.
GN R02996 OR SMC03097.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
Ox NCBI_TaxId=382;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RC MEDLINE=21396507; PubMed=11481430;
RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godle T., Goffeau A., Kahn D., Kiss E., Leleau V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsberger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galbetti F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).

```

```

DR EMBL: AL591792; CAC47575.1; -.
DR InterPro: IPR000184; Bac_surfAg_D15.
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR001412; tRNA-synL-1.
DR Pfam: PF01103; Bac_surface_Ag_1.
DR PROSITE: PS00178; AA_tRNA_LIGASE_1; UNKNOWN_1.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 618 AA; 65656 MW; E51276B829B1F78 CRC64;

Query Match 4.7%; Score 220; DB 16; Length 618;
Best Local Similarity 19.6%; Pred. No. 0.000117;
Matches 168; Conservative 108; Mismatches 279; Indels 304; Gaps 35;

Oy 117 PLISLE-----LF-----AOESTEMGINPDY-----PEYQGEOPNSEVVP 155
Db 8 PVSVAQAHAFKIFGRFPESAEEVQV-IDPRYITLTFEPGTDELEALENSQLYQD 66
Oy 156 TLEPKPLIKRLVRLFNDGVNKPRLKAPYQSSQGETSAIGSSHQKTEPVANIKAA 215
Db 67 QEEPVSGDLGAIKARDRD-----RLAVLYEKARYGTVSILVNGD----- 110
Oy 216 LEDITQESAMDLNGLIPRLROTALVAARAVGYDDLSITINSIGEVNVIHDLGEPIYI 275
Db 111 IDSLPPDAFPDGOVPVAVVYARAPFTLTGVLE-----GDAARL 152
Oy 276 DYRAVEV-RGEADDKAFTVADEPRLIGDFHNGKYEYKKNLIENASAEHGFDDGRL 334
Db 153 DPAAYDLKRGARADSTLIKAGEQI--VNDL-----KEQSRPLAKLA----- 192
Oy 335 DRSYVDILPDMTAVYSLIYDGTQYREDEVVEFTITDEKTN-----QLTTPDKLPVREL 389
Db 193 ERSYVAHATSTVYVTLRADGG-----PVAPVGNLTVSGARTVDPD-----F 234
Oy 390 LEOLITVNMGEAVNLOAVRALSNDLIATRYFNMYNTELVPPERQIQNDQVSFEQSSSR 449
Db 235 VKQYSRLNHGFRYPENIRKAEERLQNLNFFSV----- 268
Oy 450 TEPAQVDESTLEPIETVELTDGILMDISPIEFSASNLIDOKLNLVAAKARHLYDMPDR 509
Db 269 -----TINEADGLAPDGT-----IPMNIQVSEGNHFRFGCG 301
Oy 510 VLAINHDDGV-----NRSLIGISDAVSANVARAILPDESENEVIDLPERTALANKRTP 562
Db 302 V---STTDGLGLOGYMGHRLFGRAESL-----RIEGSVDRLTGETTVAGADYS 347
Oy 563 ADVYQSKK---VPLVYVFASDKPRDGOIGLGMGSDTGTRLVTKFEHNLINRDGY----- 614
Db 348 AGILPAKFGAFGPASTFTASVKA-----IYDPDAVSAKTVT 384
Oy 615 --AGAEIRLS-EDKGVKLYATKPLSHPLNDOLRATLTGYO--OEVFGHSTNGFDLSTRTL 669
Db 385 AAGAAFAFELSPEDTFV-----GAEVGMDVVDATG----- 415
Oy 670 EHEISISIIONGWNNRTYSLRYRLDKLTKTOAPETMODLPVDVNGKPSQEALLAGVAVH 729
Db 416 ---SNSYI-----TAA-----LPFEVY----- 429
Oy 730 KTVADNLVMPMRGYR-----QRYSLLEVSGSGLVSDANMALIARAGISGVYSEGDNAVSNRA 785
Db 430 RQARDKLNPTGCRYALINAKPSYEIEGKTFRSS-----FEASAGYV-----AFGTEKR 479
Oy 786 HONTGCIQAG-YIWSDFNHVYRLRFAGQDOSTIRGVAHDSLPSIDKGYLTGGVLA 844
Db 480 FVLAKGLGAGVLYVGDELSDIPATRRFELGGGSVNGYQYQELSPRDADDELTGSGSYVS 539
Oy 845 GRAEYVEEMKDLRLAVFDIGNADKGT--NDTKIGAGGVWMASSVGVQVRVADYATGV 902
Db 540 GSELRIAVYDTIGVVPFIDAQTVSDTAPDSDIRAGAGIGLRATPGPIRLDPAVL 599
Oy 903 KE--EGNPIKLHFFIGTPF 919
Db 600 NKYPGSTDYGIYAGIGSPF 618

```



```

RESULT 25
ID 09P128 PRELIMINARY: PRT: 739 AA.
AC 09P128:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Outer membrane protein.
GN CJO129C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCCT 11168.
RX MEDLINE=20150912; PubMed=10688204;
RA Parthill J., Wren B.W., Mungall K., Kelley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtrop S.,
RA Jags K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajadream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett J.B.
RA "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
DR EMBL: AL139074; CAB72613.1;
DR InterPro: IPR000184; Bac_surfag_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 739 AA; 83197 MW; B3C1891A8AA781BF CRC64;

Query Match 4.6%; Score 216.5; DB 16; Length 739;
Best Local Similarity 18.1%; Pred. No. 0.00024;
Matches 172; Conservative 142; Mismatches 317; Indels 317; Gaps 43;

QY 68 ARLNA-----GLAKPQSOALDVNFDOSPIRIGESPPGLDMVIEETTPLS 119
DB 13 AMANATATKIDIKFGLNHLNMTSAINIAGL-----KIGEEINPAKIMTALIN-----LY 61
QY 120 LEELFAOESTEMGINDPIPEYQGEOPN-SEVVVPPILEPEKPL-----IKRLYARLF 173
DB 62 KONTEINAVE---NNGLLEITIVTEKPIAVYITGIASNDRKQVESILGKR--GTL 116
QY 174 NDGVNK--VPRKAKFYOSS-----OSGFSALGSSH----- 203
DB 117 DEGNKKAERIKAYEAKSYFDITVEYKKKTLNMDGLEFYNRGENIITIDNVHLSG 176
QY 204 OKTEPYANIKAALEDITGE-----SANDLNGSIPRLROTALVAARAGYDI 250
DB 177 AKKFSYSIEPVAIVNKRKEFGMMWRNDGKLKVFELSDSSRIADEVM-----KKYLDV 232
QY 251 DLS-----IIRNSIGEDVVIHIDGEPYIDYRAVEVGEADKAFITTVADENPLIG 304
DB 233 QVSSSYLTQYTDYQANLYTFIKK--GKPYKI--KSIISINPLFDDKQNAQTVKDLSSAG 289
QY 305 DVFHGKYEYKKNLLENASAEHGYFDGRMLDRSVYILPDNTADVSLIYDTGYRFDVY 364
DB 290 KTNIEDIRKDKTIEFGSADIGY---AFVEYYPDIQKMDQTOEATVFKV---IPHDVY 343
QY 365 VFTIDPTNLTTPDKLPVKRELLEDLLTVNMGAVNLAVALNSDLIATRYFNMYN 424
DB 344 YIRNVIIISGNSRTVD---RVIRRELY-----ITEGNLVNRDTLSSKNALKATKTSFDDVN 395
QY 425 TEIYPEREQIQNDQVSPFQSSSSTPEPAQVDESLPIETIEVEL-----TD 471
DB 396 -----IKEEKYDDTHIDILVYKKEASTGAISGIGYSSD 430
QY 472 GILMDISPIEFSASNLIDKLNLVAKARHLVMDPDRLVLAINHDDGVNRSITLGRISDAV 531
DB 431 GILNLS---LSDTNIFPSGI-----KSSVSVDKSD-----TLSGRIS--- 466

```

```

QY 532 SAVARAILPDESENEVIDLPERTALANKRTPADVYQSKKVLVYFVADSKPR--DGIGL 589
DB 467 -----LVN-----PRVDSQYSL 479
QY 590 GMSDGTGRLVTFEHLNIRKDGQAGAEILRLSEDKGKVLKATKPLSHPLNDQLRATLG 649
DB 480 G-----GT-----LYS-----ND----- 487
QY 650 YQOGEVCHSTNGFDLSTRTLEHESRSIIQNGGWNRTSYRYRLDKLTQAPETWQDLR 709
DB 488 YEMDNVSEKNYGFPI---TIGROPARY-----NVSILYNLEO----- 522
QY 710 VDFVNGKPSQEALLAGVAHVKTADVNLVNPARGYRQRYSTLEVGSGIVDANNAIARAGI 769
DB 523 SDIYHLSP--LRTGYELKSIKSS--ITPAITFNDDVYLRSGII--ASTSLFVAGL 577
QY 770 SGVYSEGDNAVGSNRAHOMTGIO--AGY-----INSDENHVPYRLRFPAGG 815
DB 578 GGDQEF---ISSSKFNFYQGLQDYGYDLIYRKASFYKVDQGY--LPINQRYLGG 631
QY 816 DOSIRGAHDSLSPIDSKGYLTGGQVLAAGTAENYEFMDLR--AVFEDIGNADYKGF 873
DB 632 IKSIRGESKTVSPKQMGDEIGGTIAFANSVELSPILDIRIKLGSVFEDYGMIGRKNL 691
QY 874 TNDTKIGAGVVRWASPQGVRVAVATGV--KEGNPIKLHFIPTPF 919
DB 692 DELKRMSTGIGIMITPPIGLQLVFAKPLNDKKGDDTINSEFNLGTRF 739

```

```

RESULT 26
ID 09A7W6 PRELIMINARY: PRT: 628 AA.
AC 09A7W6:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein CC1603.
GN CC1603.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15.
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uitterback T., Tran K., Wolf A., Yamathvean J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005835; AAK23582.1;
DR TIGR: CC1603;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 628 AA; 68203 MW; 92DAD1722C682295 CRC64;

```

```

Query Match 4.6%; Score 216; DB 16; Length 628;
Best Local Similarity 21.8%; Pred. No. 0.0002;
Matches 110; Conservative 87; Mismatches 236; Indels 72; Gaps 17;

QY 456 DESTLEPIETVELDGLMDISPIEFSASNLIDKLN-----VAAKARHLV-----D 504
DB 155 DEGVQRQAVAAARLIEGEGRSADYVAGCRIVAGYADVAAEPREYVVDHADRIV 214
QY 505 MPDDVLA-----INHDDV-----NRSLIGRISDAVSAVARAILPDESENEVIDLPERT 554
DB 215 RPTFRIMGELVRLGVVDVYVTKGRTPNPEVWGRLAPVVA-----GGVYPPEDVA 262
QY 555 ALANKRTPADVYQSKV-----PLVYFVADSKPRDQIGLGMGSDGTGRLV 600

```

```

Db 263 ELERRLDRTAVYDSSVSLAGTDKASAGYRPPVVTLSDDRARTIELGAGSTSEGACVD 322
Oy 601 TKFEHNLINIDGVOAGAELELSEDKKGVKLYATKPLSHPLNDOLRATLGYOOEYFGHSTN 660
Db 323 ARWR--YNNQKADTTTVALRFAKLQRLGAEISLPHMWRPQ--QTLKINSVFRNDTD 378
Oy 661 GFDLSTRTLEHISRSITIQNGMNRRTSLRYRLDKLTKTOAPPEWMDLPDVFNGKPSQ 720
Db 379 AYNETGATVGVDLTR--RQTTAVRTTCGVSFDSLQTEQVNRN-----GLIAGRKLNL 429
Oy 721 ALLAGVAVHK--TVAADLVNPMRGYRORYSLEVSSGLVDANMAIAR--AGISGVYSFGDN 778
Db 430 ATLAGLAAYAMWDSDDLDPKRCGRMLETRA--PTYVAGDTSVPLKLAGGSAYL----- 483
Oy 779 AYGSNRAHOMGTGIOAGYIMSDNFNHPYRLRFRFAGDQSIKGYAHDSLPSIDSKGYLTG 838
Db 484 PFKQDSTVLAARVKLGAIGLGLDVPASRRFSGGSGVSGYAYQALIGPRLSDNTPQ 543
Oy 839 GOVLAVGTAEYNYEFMKDLRLAVFGDIG--NAVDCGFTNTKIGAGYGVWMAAPVGOVRV 896
Db 544 GISLVETSEFVRQKIDRMSCGAVFVDAAGATGHTETPQREDFRAGAGLGVARTDLGFGPIRA 603
Oy 897 DVATGV--KEGAP--IKLHFFIGTGF 919
Db 604 DIAAPLGRKCDPKFOIYLSIGSF 628

RESULT 27
O9JXB7 PRELIMINARY: PRT: 635 AA.
AC O9JXB7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein NMB2134.
GN NMB2134.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gunn M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecio A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Utechtack T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappunli R., Venter J.C.,
RA *Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.
RL Science 287:1809-1815(2000).
DR EMBL: AE002561; AAF42442.1;
DR TIGR: NMB2134;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 635 AA: 69801 MW: 28371F4A7A9DA98 CRC64;

Query Match 4.58; Score 215; DB 16; Length 635;
Best Local Similarity 22.18; Pred. No. 0.00023;
Matches 132; Conservative 73; Mismatches 225; Indels 166; Gaps 20;
Oy 462 PVITVELTJGIIIMDISPIESASNLQDKNLVAAKARHLIYMPDRVLAINHDDGYNR 521
Db 68 POTESVALKRFPPLVLDITQDSEIKDWEHMLPLITQOOEEVLDEQTFLEAEAPDNVKT 127
Oy 522 SI--LGRISDAVSA-----VARAILPDSENEVIDLPERITALAN 558
Db 128 MLRSKGVFSSKSVSLTEKDAVTVITGTGPRKIANVGAAILGLDLSGNTAEYVRNLEN 187
Oy 559 RKTPA-----DYOQSK-----VPLVYVVASDKP 582

```

```

Db 188 WQQPVGSDFDQDQSWENSKTSVLGAVTRKAYPLAKIGNTOAANPDATADLVNVDGSRP 247
Oy 583 RDQIGLGWGSQDGTGR-----LVTKFEHNLINRGVOAGAELEL- 619
Db 248 -----IAGDEPITTOYRPEQIYSGLARFOPGMPYDLDLDFQOAL- EONGHYSGASVQ 302
Oy 620 ----IASEDKKGVKLYATKPLSHPLNDOLRAT-----LG-----YOQEVFGHSTNG 661
Db 303 ADFRLQGDVRPVKVSYTEVRKRLKETGIRLDSREYGLGKRIAYNYLNFKNKIGSYVMD 362
Oy 662 FDLSTRTLEHISRSITIQNGMNRRTSLRYRLDKLTKTOAPPEWMDLPDVFNGKPS 718
Db 363 MDKRETTLAGISQPRNRYRNWYTSNYSNRTQNLKRAFSGGV-----YVRDRAG 416
Oy 719 QEALLA-----GVAV-----HKT-----VADNLVNPARGY-----RQRY 747
Db 417 IDARLGAEFLEAGRKIPGSVADLGNHATMLTSMKROLNNVLPBNGHYLDQKIGTTL 476
Oy 748 SLEVSSGLVSDANMAIARAGISGVYSGDNAYGNSRAHOMGTGI-----QAGYIWSDFNF 803
Db 477 GTFLSSTALIRTS-----ARAG-----YEFPPENKKLGTFLIRGAGYTVARDNA 521
Oy 804 HVPYRLRFRFAGDQSIKGYAHDSLPSIDSKGYLTGGOVLAVGTAEYNYEFMKDLRLAVFG 863
Db 522 DVPSGLMFRSGGASVYRGYELDSTGLAGPNSVLPERALVGLSEYOLPPTRTLGAVERH 581
Oy 864 DIGNAVDCGFTNTKIGAGYGVWMAAPVGOVRVNAVGVKEEKNPKLHFFIGTGF 919
Db 582 DMGDAANFRKMKLKHGSGGLGVWMSPLAFPSFDIAYGHSK--KIRWHISLGTRF 635

RESULT 28
O9PEI2 PRELIMINARY: PRT: 784 AA.
AC O9PEI2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Outer membrane antigen.
GN XF1046.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX Xylella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Artuda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Bata G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
RA Coultano N.B., Colombo C., Costa F.F., Costa M.C.R., Costa C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitaajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madelira A.M.B.N., Madelira H.M.F., Martino C.L.,
RA Marquês C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,

```

RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Melandis J., Setubal J.C.,
RT "The genome sequence of the plant pathogen *Xylella fastidiosa*."
RL Nature 406:151-159(2000).
DR EMBL: AE003941; AAF83856.1; -
DR Interpro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac-surface.Ag; 1.
KW Complete proteome.
SQ SEQUENCE 784 AA; 87991 MW; 719331DD81E8192 CRC64;

Query Match 4.5%; Score 214; DB 16; Length 784;
Best Local Similarity 20.5%; Pred. No. 0.00088;
Matches 177; Conservative 120; Mismatches 300; Indels 266; Gaps 45;

OY 228 NSGIPRLQRTALVAARAVGYDIDLSIIRNSIGEVDYIHHDLGEPYIDYRAVEVREGA 287
DB 17 NFSLPVLTQAA-----ESFVANDI-RVDG-LQRIASGVITYLLPVP-NRDTYV 60
OY 288 DDKAFTTVADDEVPLIGDVFHHG-----KYETK 315
DB 61 DD---AKVADAIKALYRTGFEFENVRIDROGNIIVYKKEPRAINKLITGNKDIKSEEL 117
OY 316 KNLIEAASHEGYDGRMLDRSDVLLPDTADVSLIYDTGQYRDEVEVFTIDPKTNQ 375
DB 118 KGLSEIGLSEGGTFDRSLDRV-----TQELKROYNNRGKYNV-----Q 156
OY 376 LTTDPDLKPYKRELBOLLTNNNGEAYNLQAVRALSNDLATRYFNNVNTIEVPEREQ- 434
DB 157 MTT--TTTPLDRNRVDYTLAIKKGAKAKIRHI-----NLGTGKFNKKDKMSMESKEHN 209
OY 435 -----IQNDQVS-----FEQSSSRTEPAQVD-----ESTLEP---VIEFTVELTDG 472
DB 210 WASWYRRDDQYSKEKLSGDEKLSNWTYLDRGVDFNIDSTQVSIPEKHMMFTAGVTEG 269
OY 473 ILMDISPIEFESANLI-ODKL-NLVAAKARHLDMDDRYLATINHDDGVARSLTGRSDA 530
DB 270 DQYKISSIVTGTVPQKTEKLVIPKTDGIDF---SRFL-LEYSASAIINLNSIGYA 324
OY 531 VSAVARAILPDESENEVIDLPERTALANRKTTPADVYQSKKVPV--VYFVASDPRDQIG 588
DB 325 FSKVNN--PIPTANRA-----DRTVAVN--LHVIPGRVTVYRQILFKKNTTSD---- 368
OY 589 LGMGSTGTGLYTKFEHN-----LINRDGYQACAEIRL-----SEDKKGVKLY 631
DB 369 -----EVLRRERKOFENSWTQQAIDRSKIRLQRLGYFEAVDESTPVPQSNQVDI-VY 422
OY 632 ATKPLSHPLNDQRLATLGGQEVFHHST-----NGFDLSTRLEHHSITQ----- 679
DB 423 TVKETT--SGSFQVGLGT-SKTYGVTTSVQLSÖNNFLGSGNRSVDSRSRYQDRYSFS 478
OY 680 -----GNMRTY-----SLRYRLDKLKTQ--APPETMODLIVDFVNGK 716
DB 479 YNPFPTDNGVSLGYMLAQOKLDYSPFMAQVYSKRMSQGTITGCIPTEND-TVSWVIGA 537
OY 717 PSOEALLAGVAHVKTIVAD-----NLVPMKGYRQRYSLV 751
DB 538 DSNQITTFPSPKALIIDYDAVGQRTFPAWPTELGMARDTRNDYDPMPLGMQRIQAEV 597
OY 752 ---GSS-----GLYSDANMAIARAIGSYVSPEDNMGVSKRAQOMTG 790
DB 598 TLPGSTIKYKKINYOISKWPIIPALVLTNRLEV-----GYDDD-YGKSHRTILPD 647
OY 791 G1OAGYIWSDNFNHVPYRIYRFAGGDSIRGYAHDLSLPSD-----KGYLNGGVLA 844
DB 648 G1VVA-----TASGLPFFENFYAGTNSVGRFNDNTLGRSEVTALYNQÖPILGGSFKYV 701
OY 845 G1AEVNEEFMKD---LRLAVFGDIGNAYD--KGF-TNDTKIGAGVGRVMAAPGVQVVDV 898
DB 702 GSTEMFFPKLFDSPSARISAFIDFGVFNQVNVNFKANELIRASGVALLMRAPIGISISV 761
OY 899 ATGVKEEGNP--IKLHFFGTPE 919
DB 762 APPIKKNENDEIERLOFTFGQF 784

RESULT 29
O8ZH58 PRELIMINARY; PRT; 795 AA.
AC O8ZH58;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative surface antigen.
GN YP01052.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxId=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE-21470413; PubMed-11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Tlhalil R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holtroyd S., Jørgensen K., Kariyasev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.,
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
DR EMBL: A414146; CAC89894.1; -
DR Interpro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac-surface.Ag; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 795 AA; 87838 MW; 8C4581328D38F821 CRC64;

Query Match 4.5%; Score 211.5; DB 16; Length 795;
Best Local Similarity 19.1%; Pred. No. 0.00055;
Matches 172; Conservative 139; Mismatches 306; Indels 283; Gaps 42;

OY 148 NSEVWVPPTLEPKPG-LIKRLVA-----RLFNDGVNKPRLKAKFYQSQSGERSAI 199
DB 45 NMFVVGDIYSDDDIGKTALPATGNEFDYRVLRDGNLTIVQK-----EKPTI 94
OY 200 GSSHOKTEPYANIKAALEDITQESANDLNGSIPRLRQTLVAARAVGYDIDLSIIRN-- 257
DB 95 ASI-----TESGKAKVAKEDMLQ--NLEASGYRVGEA-----LDRFTTISNIE 134
OY 258 -----STGEVDYIHHDLGEPV---YIDYRAVEVREGA-----DKAFTT----- 294
DB 135 KGLEDEYYSYKYSASAKAVVTPLRNRVDLKLVTFTGVSAKIQOINIVGNHSFTDELI 194
OY 295 ---VADEVPL--LIGDVFHHGKYEKKNLIEAASHEGYDGRMLDRSDVLLPONTAD 348
DB 195 SRQLRDEVFMWVVGQ---KKYÖKOKLAGDLETLSRYLDGIGYARFNIDSQVSLTPD 250
OY 349 VSLIYDT-----GTOYRFDEVVFTIDPKTNQLTTPDKLPVARELLEQLLTVMGEAYN 403
DB 251 KKGIYVTINITEGPOPKNSVIY-----SGNLAGHÖSEAEKLTKEPGLFN 297
OY 404 LQAVRALNDL--IATRYFNNVNTIEVPEREQIQNDQVSFEQSSSRTEPAQVDESTLE 461
DB 298 GSKVYTMEDDIKMKLGRGYAVRVVTOPE---IND-----DDKTVK 336
OY 462 PVLETVELDGLIMDISPIEFESANLIÖOKLNLVAAKARHLVD--MPDDRYLATINHDDGV 519
DB 337 ---LHINVDAGNRFYVRHRLHFEBSNDTSKD--SVLRERMRQMGASLGNQDVE--GKERL 389
OY 520 NRSILIRISDAVASAVARAILPDESENEVIDLPERTALANRKTTPADVYQSKKVPV----LVY 575
DB 390 NR--LQYF-----ETVDY-----EQGRVPGADIVD 413
OY 576 FVASDKPRDG---QIGLGMGSDGTGLYTKFEHNLIINRDGYQAGAEIRLSEDKKGVKLYA 632
DB 414 VYKVKERNYTGSLNFGIGYGTESG---VSFQYQÖDMMWLTGNTGVGINGTKNDYQTYA 469

```

OY 633 TKPLSHPLNDQLRATLTGVOEVEFGHSTNGFDLSTRLEHEISRSIIIONGWM-NRTYSRLY 691
Db 470 EFTLMDDY-----FTYDGVSLGRIRFYNDKADNADISGTTNSYGC--- 510
OY 692 RLDKLTQAPPEYTMODLPVDFVNGKPSOE--ALLAGVA-VHKTAVDNL-----V 737
Db 511 -----ADGTLGFPINENNLSLRVGVGYVHNDSDMLPQVAMMYLESV 552
OY 738 NPMGYROR-----YSLEVG-----SSGLVSDANMAIARAGISGVY---SF 775
Db 553 GERGVYGRGEGFTTDDFTLNLGWTYNNLDRGFPTSGVSKSVNKKITVPDGNDFEYVTF 612
OY 776 GDNAV---GSNRHQMGTGIIAGYIWSDNFNHVPYRLRFAGGDSIRGYAHDLSF--- 829
Db 613 DTSAYOPLNDRSMVLLGRRLGIGDGSKEPMFENFYAGSGSTYRGFRSNNGPKAA 672
OY 830 -----ISDKYLGGVAVGTAE-----YNEFKMDLRLAVFGDIGNAYDKGFT 874
Db 673 YVANGATVNSTDAVGNNMAVAISIELIPTPFISSEKYSNSVTSIFIDSGTWTDTNME 732
OY 875 NDKR-----IGAGVGYRASPYGVAVDVATGYKE-EGNPT-KLHFFIG 916
Db 733 NTAKTRAGIPDYGKASINIRVSAGVALQWMSPLGPIVFSYAKPKYKDEGDKSEQFOFNIG 792

RESULT 30
OBYE13 PRELIMINARY: PRT; 623 AA.
AC 08YE13:
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DE Outer membrane protein.
OS BME11895.
OC Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX NCBI_TaxID:29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE-20020109; PubMed-11756688;
RA Delvecchio V.G., Kapralov V., Redkar R.J., Patra G., Muijer C., Los T.,
RA Ivanova N., Anderson I., Bhatnagar A., Lykdis A., Renik G.,
RA Jablonki L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Teleson J.-J.,
RA Haashekorn R., Kyprides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL: AE009623; AAL53076.1;
KW Complete proteome.
SQ SEQUENCE 623 AA; 67227 MW; 07C43D0F6ADBF8FE CRC64;

Query Match 4.4%; Score 209; DB 16; Length 623;
Best Local Similarity 21.0%; Pred. No. 0.00051;
Matches 146; Conservative 79; Mismatches 265; Indels 208; Gaps 28;

OY 310 GKYERKKULIENASA-----EHG-----YFDGRW-----LD 335
Db 42 GKEADLKSVIEGASGLVSDADKPAASGASGLLAKAGDYRLILSALYGCGRGTTISIKVD 101
OY 336 -RSVDVILPD---NTADVSLIYDGTQYRFDEYVFTIDKTNOLJTTPDKLPVKR- 388
Db 102 GREANDIPDPEIPNNNAVATVDPGQPLFSRAISNIAPPGRNR---DKVTPPEAG 158
OY 389 -----LLEQLTVMNGEAYNQAARALSNDLIARIFVMVTEIYFPEREIQ 436
Db 159 FAPGGEAKSCTILAEKRLAVAMROEGAKARVGEDVAVADHADRVSADIALDGRKAY 218
OY 437 NDQVSSFOSSSSRTEPAQVDEST-LEPIYETVELTDLGILMDISPIEFSSAMLDKLNLY 495

```

```

Db 219 YGPVSV---VGTRAMDPOFVAMMTGLKPCQE----- 246
OY 496 AAKARHLVMDPDDRYLAINHDDGVNRSLIGRISDAVSAVARAILPDESENEVIDLPERTA 555
Db 247 -----YD-PDDIENM-----KKRLGRME-----VFRAAMFELEADK----- 275
OY 556 LANRKTADYQSKKVPVLYFVASDKPRDQIGLQMSDGTSLVTKFEH-NLINRDQYQ 614
Db 276 -----IEPDGSLPTILNVQERKPRRGFGAEYSTIDGFGVTSYMHRNLLRG---- 322
OY 615 AGAELRLSEDKGY-----KLVA-----TKPLSHPLNDQLRATLTGVOEVEFGH 657
Db 323 -GERLRPDADAVSGSGSDNSFPKNTTYLLGASFAKPGVYTPPTDVFATLDAREV--- 378
OY 658 STNGFDLSTRLEHEISRSIIIONGNMRTYSLRYRLDKLTQAPPEYTMODLPVDFVNGK 717
Db 379 -----LDAYET-----SINAKGTGFIQS-----DELSGALYANNSQGHFVDVFGK- 421
OY 718 SQEALLAGVAVHKTAVDNLVNPARGYRQRYSLVEGSSGLVSDANMAIARAGISGVYSE-- 775
Db 422 -RDFTTAG-----LEGNL-----YDSRNKKPDPSSGFYLVGN-----IQPYEFHY 462
OY 776 GD-----NAYGSNRHQMGTGIIAGYIWSDNFNHVPYRLRFAGGDSIRGYA 823
Db 463 GNFAITPRTAGRTYHGFQIDRYVLGRLKVGSIIGSADLPSSQLFLAGGGGVNRYG 522
OY 824 HDLSPLSDKGYLTGGQVLAFTAEYNEPMKDLRLAVFGDIGNAYDKG---PTNPTKIG 880
Db 523 YRNIGVSAGNGNIIIGRSLVEANGEVTRITDSIGNAVAFADAYVEKSPDSEDMRVG 582
OY 881 AGGVVMASSVGVQVAVDVATGYKEEGNPKLHFFIG 916
Db 583 VGGGLRYLTSLGPIRLDVAVPLNRRSGDPNRYGYG 618

RESULT 31
O25369 PRELIMINARY: PRT; 916 AA.
AC 025369;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Protective surface antigen D15.
OS Hp0655.
OC Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-26695 / ATCC 700392;
RX MEDLINE-97394467; PubMed-9252185;
RA Tomb J.-F., White O., Kervatave A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S.,
RA Lotous B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Wattey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RT Nature 388:539-547(1997).
DR EMBL: AE000579; AAD07720.1;
DR TIGR: HP0655;
DR InterPro: IPR000184; Bac_surfaq_D15.
DR Pfam: PF01103; Bac_surface_Ag_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 916 AA; 102671 MW; 186D8E301CD9408A CRC64;

Query Match 4.4%; Score 206.5; DB 16; Length 916;
Best Local Similarity 19.3%; Pred. No. 0.0014;

```



```

Db      505  GSYKLHISGVKTSISLQYDKM--EDLSHEIDYLLKRIOLASPPSSNSIFLEMDGKLITSA--561
Qy      703  ETWQDLLPVDVFNKGRSQEALLAGAAVHKTVADNLVNPVRGYRQRYSLEVGSSGLVSDANN762
Db      562  -----IGHITTYDDNDNKIVPRNGYL-----582
Qy      763  AIAKAGISGVSYSF--GDNAVGSNRH-----QMTGGIQAG-----YI797
Db      583  -----VSGTQFPAGVSGDNKTYIKHEIECKRYKSFINNKTYLKLSAAGMDAGLGGKAMVR636
Qy      798  WSDNNNHVPLRLRFAGDQDOSTRGYAHDSLSP---ISDKGYLGGQYLAVGTAEYNY--851
Db      637  ISDRENL-----GDYSILRGFASGSGVGPREKKNTEG--LQGERRYTFYFSELNEPTP684
Qy      852  ---EEMKDLRLAVEGDIGNAY-----DKGFINDRIKIG--GVGVRMASPVQGVQR895
Db      685  VPDEE--NFTGAVFLDLSGVNGVGLNKKQYTPMGFNTDQSLRASVSGFGFIWTRAPAIR742
Qy      896  VDVATGVKEE905
Db      743  MDMGFPPIKK752

```

RESULT	33
092367	
ID	092367
AC	PRELIMINARY;
DT	01-DEC-2001 (TREMBLrel_19, Created)
DT	01-DEC-2001 (TREMBLrel_19, Last sequence update)
DT	01-JUN-2002 (TREMBLrel_21, Last annotation update)
DE	Outer membrane protein compl.
GN	OMP1 OR RC0202.
OS	Rickettsia conorii.
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC	Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX	NCBI_TaxID=781;
RM	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=MALISH 7;
RX	MEDLINE=21442074; PubMed=11557893;
RA	Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA	Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA	Raoult D.;
RT	"Mechanisms of evolution in Rickettsia conorii and R. prowazekii."
RL	Science 293: 2093-2098(2001).
DR	EMBL: AE008587; AA02740.1; "
DR	InterPro: IPRO00184; Bac_sufAq_D15.
DR	Pfam: PF01103; Bac_surface_Ag; 1.
DR	Complete proteome.
QC	SEQUENCE 768 AA; 86881 MW; F0327FC56408E36A CRC64;

Query Match	4.3%	Score	202.5	DB	16	length	768
Best Local Similarity	20.1%	Pred. No.	0.0017				
Matches	172	Conservative	122	Mismatches	301	Indels	261
						Gaps	43

```

0Y      164 LKRLVYA-----RLENG-----VNKPRLKAFYSSOGSGENSAI-----GS 203
Db      64 VIKRLVATSLFRNINMHITNDGNLIIVATLEPRLSSVYFSGNSKIKTNMLAKLEYTMSGE 123
0Y      202 SHOKTEEPANIKKALEDITDIESAMDLSGSIPELROTALVAAARAGVYDIDLSIIRNSIGE 261
Db      124 SLSQAKLELDQKLTLE-IYKRSGRFATTVPKIE-----NLENNKV-- 163
0Y      262 VDVYIHLV--GEPYIDYRAVEVYEGEGADKAFYTVADVP-----LLIDGVFHGKYE 314
Db      164 --KVIPIAIEGPKTGINVITYFSGNENYSOSELSIVLTKSRRMFRPLESDTYDIDPRMEY 221
0Y      315 KKNLEIASAEHGCFDORMLDRSVDVLLPNTAD-VSLII--DTGTQYRDEVAVFTIDP 371
Db      222 DKELLRFQYVSGADRVLTISASAEI---NNTKEYFLVTISIEEGEYFSGNT---TIDN 275
0Y      372 KTNQUTTDIPDKLPVYRELLBOLLTVNNGEAVYINQAVALSNDIAIATRYFN-----MYNTE 426

```

Db	276	KLTLNIDIP-----LNLTVNJKCKQVYNNMKVVDIAKI--GEFTANGPAVN--	322
Qy	427	IYFPEREOIQNDQVSFEOSSSSTPEPAOVDESTLEPIETVELTDGILMDISPIEFSASN	486
Db	323	-VYPN-----IMKNANH-----ADIKFIIEKADK	346
Qy	487	LJODKLNLYAAKARLHYMPDPRVLAINHDDQVARSILGRISDAVSAVARAILPDESENE	546
Db	347	VYKINKINI--NNLKTEDPVIRREKIEGGIINNSYTEK-----GEKNLR	390
Qy	547	VIDPERTALANKRTPA--DVYQSKKVPPLYVVASDKPRDQIGLGWGSIDGTRLVYKF--	603
Db	391	NIDYFEKVALISLAPTKADKXD-----VAVEDENKSTSSIGFDLGYNAGGLFGFRFSF	443
Qy	604	-EHNLI-----NRDQYAG-----AELRLSEDK-----KGKVIYATK	634
Db	444	LEBNLVGTGKGLLNAGVOYQSKMTTSYGGITAPHFLDRDLRISGVNTFPNNYGRGASVLTNT	503
Qy	635	PLSHPLND-QLRATLGXY-QEYFGSTNGPFLSTRLEHSRSIIIONGGMNRTYSLRYR	652
Db	504	DOSYVLHSTIGVXTSIGVEYIKEDLGEH-----IDYLLKRDILSLAPTPSSSTFLKEQ	553
Qy	693	IDKLTOAPPEPTMODLPVDYFNKGRPSQEPALLAGVAVHKTYADNLVNMRRGYRQRYSLKVG	752
Db	554	MGRFTTSA-----IGHTITYDQTDNKIYVKNKILYVSGTOEFA	590
Qy	753	SSGLVSDANMAIARAGISGVY--SFGDN-----AVGSMRAHQMTGGIOAGYI--WSDN	801
Db	591	GVG-----GNNKYIKHEVDGKYYKKSFINHKLTLKLSASGNN-----ITG--LGKILIRISD	640
Qy	802	FNHVYVRLRFFAGGQOSIRGYAHDSLSP-----ISDKGYLTLGGVYLTAETAYN-----E	852
Db	641	FNL-----GQYSLGRFANGGVGPREKVTNEG--LGGERYYKFSTELNPEPTPIPEE	688
Qy	853	FMKDRLAVFPDGIGNA-----YDKGFTYDTRKIGI--GVGYWMAQSPVGOVAVDA	899
Db	689	F--NLTGIVFMDLDSGVGVGLNKKQYQTPNGVFIYMDTSLRASSIGCFGIWYTRFAPJMDMG	746
Qy	900	TGVKEGEGNPIRLHFFI	915
Db	747	FPKKKKVDDTQHFHL	762

```

RESULT 34
09X3V8
ID 09X3V8 PRELIMINARY; PRT; 738 AA.
AC 09X3V8;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical 81.1 kDa protein.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
NCBI_TaxID=542;
RX
RN
RM
SEQUENCE FROM N.A.
RC STRAIN-ZM4;
RP
RR
RT Lee H.J., Kang H.S.;
RL "Sequence analysis of 42C11 fosmid clone of Zymomonas mobilis ZM4."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF088896; AAD2153.1; -.
DR InterPro; IPR000184; Bac_surf4g_D15.
DR Pfam; PF01103; Bac_surf4g_Ag; 1.
KW Hypothetical protein.
SQ
SEQUENCE 738 AA: 81140 MW: F169545D09168114 CRC64;

```

Query Match	4.2%	Score 198;	DB 2;	Length 738;
Best Local Similarity	20.6%;	Pred. No. 0.003;		
Matches 176;	Conservative 122;	Mismatches 344;	Indels 212;	Gaps 45
QY	127	ESTMGKINPNDYI----	PEYQGEQNSEVAVVPPTLEPEKPGITIKRLYARLFNDGVNKKYPR	182
:	:	:	:	:
:	:	:	:	:

Db 36 DSGNKAASNPADIPFDPSPFDHLP---TLSQPSIEASKA-----DGS5HLP- 79
QY 183 LKAFYSSOGSETS-----AIGSSHQTEPYA-NIK-----AALEDI 219
Db 80 --LRTSESSOGSENGRGVAKALDETETVSDPKNLLANSDBEKTAKQIEETSTYLDPA 137
QY 220 TQESAMDLNGSIPRLRQTALVAARVGY-YIIDLSIRNSIGEDVYIHDGEPYIDYR 278
Db 138 LSEPLPLSHMPMNEHVYLAKESSDLRYKVDYGLSN-----IDT-----DQAFKNS 187
QY 279 AVEYRGEGADPKFTYADEVPELIGVFHHGKETKNLIENASAEHGYEDGRNLDQSV 338
Db 188 VLSNKKKABSL-----VIG-----ARATNDELINRLIRSGGYDG---KAA 228
QY 339 DVILPDNTADVSLYD--TGQYRFDEVEVFETIDPKTNQLTDPDKLPVKRELL-----E 391
Db 229 LSTPIEKGQYVRYDVNAGSIYKLGQI-----NLTGESEPLRIAVVALNLHNGD 279
QY 392 QLTVMNGEAYNLQAVALSMDLIATRYFNKNVMEIYFPEREQIQNDQVSFEQSSSRTE 451
Db 280 PILATHEQAEEN-----NILITLPHYGYPAKI--GDRITILDD---ETHGQVTL 325
QY 452 PAQV-DESTLEPIYETVELTDGILMDISPIEFASNLQDKLNLVAKARHLXD--WDD 508
Db 326 PVNAGNNGSGSI--VSNNKHIVLDKAKHS-----HIAFRREGORYDSQWVD 372
QY 509 RVLAINHDDVNRSLIRISDAVASAVARAILPDESENEVIDLPERTALANKRTPADYQS 568
Db 373 LRQAL-----AATSLFSHVSVEPIATGR--KREDGSEYDL-----DVRQG 411
QY 569 KKPPLVYFVASDKPRQOQIGLWGSDTGTRLVTKFEHNLNRDGYQAGAEIRLSDEKRGV 628
Db 412 R-----GKKHSIAYVGYGTGEG--FKAQGSWISRNYPPEGSLTFS---GI 453
QY 629 KLVATKPLSHPLND-----QLRATIGYQOEYFGHSHNGDLSITRLEHISISII 678
Db 454 LGRROOQLSLFNRNNAAGARDVYIQIGLTAGREXY--DAWNGSFS--LGSLSQST 507
QY 679 QNGGWNR--TYSLRYSRLDKLKTQAPPETWODLPYDFVNGKPSQE---ALLAGVAVKTV 732
Db 508 QL--WQKRMVYISIGAL--TOTNERSY-----DFSRSELTFTFYLLAALPGQLGIDR- 555
QY 733 ADNLVNPMSRYR--QYSLSEVG--SSGLVSDANMAIARAGISGYVSFGDNAYGSNRHQM 789
Db 556 SNNLMPPTKCYRLNMRSPETSISGLRGYVRLMFL--DASGYVPADVNV-----LA 605
QY 790 GGIQAGYIMSDNHNHYRLRFAGGDSIRGVAHDSLSPSDKGYLTGGQVLAAGTAEY 849
Db 606 GRRVVASIEGASVOELAPSRRIYAGGGSVRGYQOQLGPKDPYKDAVGGISEBELAFEV 665
QY 850 NYEFMKDLRLAVFGDIGNAVYDKGFT--NDTKIGAGVGVRRASPYGVQVRYVATGVKKE- 905
Db 666 RRAF--GNFGLVPIIDAGQYEDAIPTPHNLRFGTGLGARITYTAFFGRIDLATPIARQ 724
QY 906 GNPDKLHFFIGTPE 919
Db 725 ESRISYVISIGQAF 738

RESULT 35
QY0YUR6 PRELIMINARY; PRT; 833 AA.
AC QY0YUR6;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein A1r2269.
GN A1R2269.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxId=103690;
RN [1]
RP SEQUENCE FROM N.A.

FX MEDLINE-21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Matenabe A., Iriuguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.,
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001)
DR EMBL: AP003588; BAB73968.1;
DR InterPro: IPR000184; Bac_surftag_D15.
DR Pfam: PF01103; Bac_surface_Ag_2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 833 AA; 89633 MW; 7F52B1E52FEA1B59 CRC64;

Query Match 4.28; Score 197.5; DB 16; Length 833;
Best Local Similarity 19.78; Pred. No. 0.004;
Matches 134; Conservative 104; Mismatches 255; Indels 187; Gaps 30;

QY 371 PRTNQLTDPDKLPVKRELLBOLLTVNNGEAYNLQAVRALSNDL---IATRYFNKNVNTETI 427
Db 210 PNATPETTEP-----RLVSEVL-----VAPQSGQLTPELETQVYVIRQ- 250
QY 428 VFP-----EREQIQNDQVSFEQSSSRTEPAQVDESTLEPIYETVELTDGILMDISPIEFS 483
Db 251 --PGRTTTRSQLQEDINAIPEGTGFPSNVQASPEDTPLGVRSFLIVQRPVL---SKVEIQ 305
QY 484 ASNLQDKLNLVAKARHLNMPDDRVLAINH--DDGVNRSLIGRISDAVASAVARAI-LPD 541
Db 306 ANPGTNVPSVLPQATADEIFRAQYKILNLRDLEGI-KELTKRYQDQGYLVANVGAPQ 364
QY 542 ESENEVIDLP-----BRTALANKRTPADYQSKKVL-----YVYVASDKPRDQ- 587
Db 365 VSENGVYTLQVABGVENISVFRNKEGQDVNEGQPIRGHTQDIITREVELPQGVFN 424
QY 588 -----GLWGMSDQTR 598
Db 425 RNTVQKDLQRYFGTGLFEDVNVSLDPGTDPYKVVNVVYVRRSSGSIAGAAGISSSGLF 484
QY 599 LVTKFEHNLNRDGYQAGAEIRLSDEKRGVLYATKP-----LSHPLNDQLRATL- 648
Db 485 GTVSYQOONLNGRNOKLGAELYOLERELELDLRFPTDWIGDPYRTSYTANIFFRRSISL 544
QY 649 ---GYQGEV-----FGHSNTPDLS--TR-----TLEHESRKIIQNGGNNRYSLSLYR 692
Db 545 ITPDKDEDIRTPDGNPDNDINGODRPVYTRLGGGVFTTRPLSANPFRAEWTASAGIQY 604
QY 693 LDKLKTQAPP-----ETWODLPYDFVNGKPSQELALAGVAVHKTVADNLV 737
Db 605 --FVSTRDADAGNLRKQCAVAFDDDNKRTSEIYPLSF--SGTGEDULLVLQAGORLNNPL 661
QY 738 NPMRGYRQRYSLSEVGSSGLVSDANMAIARAGISGYVS-----FGDNAYGSNRHQM 788
Db 662 QPTSGSFLRFG--VDQSVPYGSGNIFLTR--FRGSYQOYLVPVKFTGP-----SKGPEYI 711
QY 789 TGGIQAGYIMSDNHNHYRLRFAGGDSIRGVAHDSLSPSIDKGYLTGGQVLAAGTAE 848
Db 712 AFNIQGGTVLGD---LPPYEAFTLGSGNSVGRY-----EEGALGSGRSEVQASVE 758
QY 849 YNFEFMKDLRLAVFGDIGNAVYDKGFTNDTK-----IGAGVGVRRASPYGVQVRYVATGV 899
Db 759 YRFPVSVSGALFEDVGS--DLGSTRTALEVNLKSGSGYGLGVAVQSPGLGIRIDY- 815
QY 900 TVKKEGNPIKLHFFIGTPE 919
Db 816 -GINDGDS-RINFGIGERF 833

RESULT 36
QY0ZQ48 PRELIMINARY; PRT; 776 AA.
AC QY0ZQ48;
DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TReMBLrel. 19, last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
DE Putative outer membrane transmembrane protein.
GN OMP OR R01502 OR SMC02094.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN
RP
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Bolstead P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaut P., Vandenbol M., Weidner S., Gallibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: AL591787; CAC46081.1; -
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ
SEQUENCE 776 AA: 84520 MW: CE261E6372A94726 CRC64;

Query Match 4.2%; Score 197; DB 16; Length 776;
Best Local Similarity 21.2%; Pred. No. 0.0038;
Matches 165; Conservative 114; Mismatches 315; Indels 186; Gaps 41;

OY 223 SAMDNGSIPRLROTALVAARAVGY-DIDLSIRNSIGEVYIIHD---LGEFVYIDYR 278
DB 70 SMDIDDSVRL-----YATGFSFDSVSTV---SGGSLVSVSENLQNVFVNGNR 118
OY 279 AVEERGSAGADKAFITVADVEPLIGVFIHGKETKKNLIENASAEHGFTGMDLRSV 338
DB 119 KIK-----DCKLQGVVFTQ-PL--GPYSFATVETDIOAIRVAYAIGRSD---VTVTT 165
OY 339 DVI-LPDMTADVSLIYDTGTQYRFDEVVFF-----TDPK-----TQOLT 377
DB 166 QVPIAGRVNLAFFVINEGERTKITQINFVGENYSDGRQSVATRESGIFSLTKDV 225
OY 378 TDDPKLPVKRELLBQLITVMKEAYNLQAVRALSNDLIATRYFMVTEITVFPEREQIN 437
DB 226 YNDPKRADELLEKRFY-YNRGYA-----DQVVSSEALNEATNEYT 267
OY 438 DOVSFEOSSSRTREPQOVDSLTLEPIVETVELDGLIMDSPIEFSSNLIQDKLNIVAA 497
DB 268 VTTITEGPRYPDGPVNI-ESTVEG-IDAEELR-GLVQSRGFTVYRKAD-IQSTMSISK 323
OY 498 K-ARHLVDMFDDRVLA1NHDDGVNRSILGRISDAVSAVARALP----- 540
DB 324 RVASEGY--PFAVVTFRGNLDLANHTI--AVDVLVDOGEAIVYERIEIKNTTRDYVIR 379
OY 541 ---DESENEVIDLPERTALANRTPADVYQSK-----KVLVYFVASDKPRDG 585
DB 380 REFDEVEGDAFN-QEWAARAKRLALGYFSSVNISTOPGSAADRVIVVDQDQSGSF 438
OY 586 OIGLGMSPTGRRLVTKF---EHNLLNRQYOGAGELRS---EDKKGVKYLATRP-L 636
DB 439 GIGAGTSACDGGGFLVEASIEEKNFLGRQY---IRLAAGKESQTYNVSFTPEYFL 493
OY 637 SHPLN---POLRATLGYOQEVFGHSTNGFDLSTRTLEHHSRSIIQNGGNRTYSLRYRL 693
DB 494 GYRLAAGFLFNENDFDDDNYSYNDGFSL-----RYTAPITE---NSTLTLYNY 542
OY 694 DKLKTAQAPETMODLPVD-FVNGKPSQEQALLAGVAVHKTVDNLVPMRGYRQRYSLYEV 752
DB 543 TELEVFSGDDELSS-PYDRVIDOSPWTRSSISQSTIYNTL-DDAOLPHGIGLASVYQERA 600
OY 753 SSGLVSDAMKALARAGISGVYSPGNAYSNSRAHQHTGCI--QAGYIMSDN-----FNHV 805
DB 601 GLGGSYDFYKLTGKA--KWYTYLHDEA-----DIIGSLSGSAGHLFETSGSLEVFQD- 650

OY 806 PYRLRFAGGDSIRGVAHDSLSPISDKCYLGGQVAVGTAENYERF-----KDLRLA 860
DB 651 -----FQLNSNDIRKGFERNIGLGPFRMNGDGLGTTYETFAAEATFPLGGLPRDSGFGA 704
OY 861 VFGDIGNAYDGFNTDRTIGAGVGR-----WASPGGVRYDVATGVKEE 905
DB 705 LFVAGTILYG----NDVEIRGEGESYRGDMASLRASVGSILWASFPGLRDYAVPAVKE 760

RESULT 37
09ZLH9
ID 09ZLH9 PRELIMINARY; PRT; 906 AA.
AC 09ZLH9:
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, last annotation update)
DE Putative outer membrane protein.
GN JHP0600.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN
RP
RC
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Noir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Carnuso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 387:176-180(1999).
DR EMBL: AE001492; AAD06181.1; -
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ
SEQUENCE 906 AA: 101438 MW: DE81214EF452CD29 CRC64;

Query Match 4.1%; Score 194.5; DB 16; Length 906;
Best Local Similarity 19.9%; Pred. No. 0.0069;
Matches 192; Conservative 131; Mismatches 311; Indels 333; Gaps 49;

OY 116 TPLSLELFAQEST-----EMGINPNDYIP-ETV-----GEOPNSEVVY 153
DB 37 TPKEAQKNEAONETSOSNOTPKKEMKYSISYVGLSMDMLANEIAKIRGDMVDSKID 96
OY 154 PPTLEPEKPGILIRLYARLFDNGVNVPRLKAKEFYSSQSG--ETSAIGSSHQKTEPYAN 211
DB 97 TAVLALFNQGYFDVYA-TFENGI-----LEPHDEKARIAVEIKGYTEKEDGAKSQ 150
OY 212 IKAALBDITQESAMDNGSIPRLROTAL-VAARAVGYDIDLIIIRNSIGEVYIIHDLG 270
DB 151 MGIKKGDTPEQKLE-----HAKTALKTALEGGGY-----GSV----- 184
OY 271 EPIVYIDRAVEYRGEADKAFITVAD--EVPLLIGDVFNHGKYEYTKKNLIENASAEH- 326
DB 185 -----VEVTERKVSSEGLLIVFDVNRGDSIYIKQSIYESGSDKLRVRVIESLANKQ 225
OY 327 -----GYFDR-----MLDSVD--VILPP-NPADVSLYID 354
DB 236 RDMGMMGLNDKLLDQLEYDSLRIDQVYMRGLDHISSPFLKTFSTHDAKLHYK 295
OY 355 T-GTQYRDEVEVFTIDETKNQLTDPDKLPVKRELLBQLITVNNGEAYNOAVRALS 412
DB 296 VKEGIDYRISDLI-----IETDNVVPFLK-TLKALKVKKKDVFNIEHLPADQ 343
OY 413 DL---IATRYFNKVNTEIVFEREQIONOVSEFOSSSSRTPEAOVDEST-LEPVITVE 468
DB 344 ILKTEIADKGYAFA--VYKPDLDK-----DEKNGLVKVIYRIE 379
OY 469 LNDGLIMDSPIEFSSNLIQDKLNIVAAKAKHLIDMPDR---VLA1NHDDGVNRSILG 525


```

Db      380 VGD--WVHNDVYISGORTSDRI-----IRRELLGPKKYNLTJKIRNSENSILRR--LG 430
OY      526 RISDAVSAVARAILPDESENEVIDLPERALANRKTADYQSKKYLVLVYFASDKPRDG 585
Db      431 FFSKVYIEEKRV-----NSLMDL-----LVSVEERTG 459
OY      586 QT--GLGKSDGTGRLVYFEHNLINRDGYQAGAEHLRSLD--KKGVLYATRP----- 635
Db      460 QLOFGIGYSGYGIML-----NGSVSEKNLEGTGQSMGLYANIMATGGGRSYPCMPKAGRM 515
OY      636 -----LSHPINQLRATLGQOEVEFGHSTNGDLSTRTL--EHEISRSILIONG----- 682
Db      516 FPGNLSLTNP-----RIF-----DSWYSTINLYADYRISYQYIQGGGFGVN 558
OY      683 -----NNRTY--SLYRLDKLTQAPPETMODLPVDFVN--GKPSOELLAGVAVHK----- 730
Db      559 VGMGLGNRRHVSLGYNLNTKLLGFSSPLNRYSSVNEVASPRQCSPTASVILINRLSG 618
OY      731 ---TVADNLVNP-----MGYROR-----YSLEVG-----SSGLVSDA 760
Db      619 RPLVPESSGSPGAIYTSPEIKGIMDRDYHRTPTSSFTLDVYDNDYFPRNGVIFS- 677
OY      761 NNALIAAGISGVYSPD-----NAYGSNRAHQMTG-----IQAG 795
Db      678 ---STATMSGLPSSGTLNSMNGLVNRTKYVGFAYVHLOKYLIDLIARFKTOGG 733
OY      796 YIWSDNF--NHVPYRLRFPAAGDOSIRGYAHDSLSPISDKYLGGQVLAVGTAEVNYEFM 854
Db      734 YIFRYNTDYLPLANTSTFYMGQVTVYVFRNGSITPKDEFGLMGDCIFITASTELSTGYL 793
OY      855 K--DLRLAVFGDIG-----NAYDKGFTNDT--KIGAGVGV 885
Db      794 KAAKMLAFEFDFGFLFKPTRGSEFYAPATTANPKOYGVGAGPERATWASGLQI 853
OY      886 RNASPYG 892
Db      854 EWISPMG 860

RESULT 38
O51930 PRELIMINARY: PRT: 789 AA.
ID O51930
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Outer membrane antigen OMA87.
GN OMA87.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PBA100;
RA Ruffolo G.C., Adler B.;
RL Infect. Immun. 64:0-0(0).
RN [2]
RP SEQUENCE OF 750-789 FROM N.A.
RC STRAIN=9222;
RA MEDLINE=95369730; PubMed=7642134;
RA Delamarche C., Manoha F., Behar G., Houlgatte R., Hellman U.,
RA Wroblewski H.;
RT "Characterization of the Pasteurella multocida skp and fira genes.";
RL Gene 161:39-43(1995).
DR EMBL: U60439; AAC44600.1; -
DR EMBL: X74357; CAA52399.1; -
DR InterPro: IPR00184; Bac_surfAg_D15.
DR InterPro: IPR001064; CrystalLin.
DR Pfam: PF01103; Bac_surface_Ag_1.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Hypothetical protein.
```

```

SQ      SEQUENCE 789 AA; 87437 MW; DCID004AAEB2CD9E CRC64;
Query Match 4.1%; Score 194; DB 2; Length 789;
Best Local Similarity 19.0%; Pred. No. 0.0058;
Matches 154; Conservative 125; Mismatches 304; Indels 228; Gaps 37;

OY      247 YVDIDLSIIRNSIGEVDVLIHDLGEPYIDYRAVEVBEGADKAFYTVADVPLIGDV 306
Db      69 YDDVKASREGNTL-----VTVMPKPVISN--VYIVGNKSSIPDAIKQNDANGFKVGDV 121
OY      307 FHHGKYET--KKNLLENASAEHGFDGKWLDRSDVY---LPDNTADVSLIYDTGTQYRFD 362
Db      122 LMKALKEEFRRKGIYEHVNSV-----GRY--NAKYDAIYNTLPNNSAEIKIQINEDDVALFK 175
OY      363 EYVE-----FTIDPKTNO--LTTDP-----DKLPVRELLEOLLYVMGEAYMLQA 406
Db      176 EITFEENAEASSGKLADOMELDTDSMKKLGKNGKFDQYQFNKD--LETIRSYLDRGY----- 230
OY      407 VRALSNDLIATRYFNNAVTEIVEPEREOIONDOVSFBOSSSSRTEPAQVDE----- 457
Db      231 -----AQFQLDITV-----KLSDDKKEPCLISEGDLVYTKT 263
OY      458 -----STLEPIYETVELDTGLMDISPIESASMLIODKLN-----LYAAKARH 501
Db      264 RVSGMWMGMSAEIAPLLETITOL--NGLFRRTSVLEVQQRN--KSKLGERGYATAQVAVHP 320
OY      502 LYDMPDORVLAIHNDGVNSIL-----GRISDAVSAVARAILPRES---ENEVDLP 551
Db      321 TFD--EODKTTSLDPIYBAGKSYVROIRFEGNNTSSADSTLRQEKROEGMLSELVELG 379
OY      552 -----ERTAL---ANRKTADYQSKKVPYLVFVASDKPRDQIGLGSDGTGRLVYKF- 603
Db      380 KLRIDRTGYEVSVEYTKEA--IPGSDQVDVLYKKERTGSLINGIGTGESLGYSQASIK 438
OY      604 EHNLI-----NDGYQAGAEHLRSL---DKGVKLYATKPLSHPLNDQLRATLG 650
Db      439 QDNFLMGSSISLGTRNDYGTVINLGYNEPYFTKDVSLGWNVSFEEDYSSKNSAGY 498
OY      651 QOEVF--GHSTNGFDLSRTLEHLSRSIIIONGMNRTYSL--RYRLDKLTQAP----- 702
Db      499 GRTSYGNTLTGFPVNE-----NNSYVLGVGYTNKLNIAPEYNRDL 541
OY      703 -----ETMODLPVDFVNGKPSOELLAGVAVHKTADVADNLVMPRGYQQRASLEVSS 754
Db      542 YRQSMKYNDQWTFKSHDF-----DLDFGNMYNSLNGYPTPGVARNIG 585
OY      755 GLV---SDANMAIAPRAGISGVYSPFDNAYGSNRAHQMTGIGIAGYIWSDNFN--HVPYR 808
Db      586 GRVTIPGSDNKRYKKLNAEAGFYPL-----DREHGWLSSRISASFADGFGSKRLPFY 638
OY      809 LRFPAAGDOSIRGYAHDSLSP-----IDKGYLTGGQVLAVGTAE 848
Db      639 QYTSAGGIGSLRGFAVGAIGAPNAIYRTROCPDSYCLVSSDVIGNMAVYASTELIYPTP 698
OY      849 YNTEFMKDLRLAVFGDIGNAVYDKGFTNDTK-----IGAGVWRHASP 890
Db      699 VADKKNQNSVYRSLFVDAASAVWNTRMKAEDKAKPAKLANVPYSDPSKVRASAGVALQWOSP 758
OY      891 VGVGRVDAVATGKE--EGNP1-KLHFFIGTF 919
Db      759 IGPLVSYAKPLKKYQGDIEQFQFISGTGF 789

RESULT 39
Q8YMP0 PRELIMINARY: PRT: 676 AA.
ID Q8YMP0
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical protein Alr4893.
GN Alr4893.
OS Anabaena sp. (strain PCC 7120).
```

```
OC Bacteria: Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
RX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:21595285; PubMed:11759840;
RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL: AP003597; BAB76592.1; -.
DR InterPro: IPR000977; DNA_Ligase.
DR PROSITE: PS00697; DNA_LIGASE.A1; UNKNOWN_1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 676 AA; 72411 MW; C577963037D1223 CRC64;

Query Match 4.1%; Score 193.5; DB 16; Length 676;
Best Local Similarity 20.1%; Pred. No. 0.0048;
Matches 191; Conservative 119; Mismatches 323; Indels 317; Gaps 47;

OY 13 MPVALAAYLRLMTSQAALQONNPANIINHVPADHTAINQAKGNPVLTPPEQIARLNA 72
DB 1 MRISAAALLTLTLVA-----SNATQAIAA-PHTASTPSK----- 35
OY 73 AGLNARFOSALDVNPNDDSPISRIGESPPGLDMSEETTPLELFOESTEMG 132
DB 36 -----QENNL-VIRVVEETPV-QLGAVTPP-----ET-----EVTFQFSTKSR 71
OY 133 INPNDYTPREYQGBPNSEVVVPPTLEPEKPLIKRLARLFND-----GVNKPRL 163
DB 72 VLAQ-----NSPVLPSTPRPVPPAKP--ATTVNLVYATDVQVIGATPEL 118
OY 184 K---AKFYSSOSGESAIGSSHQKTEPYANIKALEDTIQESAMDLNGSIPRLQTAIV 240
DB 119 QETIRVINYTGQGDYS-----QTQLRDVAALIE-----TGLE 152
OY 241 AARAVGYUDDLSIINSIGENVYIHDGEPYIDYRAVENEGADDAKFT--TYADEV 299
DB 153 AANANV-----SKTTPSGGLNVVO--VQPVV--RSLQLGCA--KALVSYVAQ-- 194
OY 300 PLLIGDVFNHGKYEYTKKNL-----TENASAEHGFYDGRML--DRGVNVLDPNTADVSL 351
DB 195 PRFGSDI---GKRPISPEGLKQAVAOVMYADNGVNLAVLSLEPRQSLININVAE--GL 250
OY 352 IYDTGTQYFDEVEFTIDPKTNQL--TTDPKLPVKKRELLQQLLTVMNGEAYNLQAVRA 409
DB 251 VSD--IKRFYVNDGDKTIDSNNGNPVGGRTKP-----FLRQQLKLGQGVFOENTVQ 301
OY 410 LSNDLATRYFNMVTEIYFPEREQIONQVSFEQSSSRTPAQVDESTLEPVLETVEL 469
DB 302 DVQQLYRGLFGOSV----- 316
OY 470 TDCILMDISPIEFSASNLIDKLNVAAKARHLVMDPDRVLAINHDDGVNSRI--LGRI 527
DB 317 -----VAFAGD-----ATKLDIYELKENGARAINLGSSYNQDVGMLGT 356
OY 528 SDASAVARAILPDESENEVIDLPERTALANRKTADVQSKRVPLVYFVASDKPRDQI 587
DB 357 N-----YQDQNTGGKNDTLAN-----V 374
OY 588 GLGMSDGTGRLYTKF--EHNILNRD--GYQAGAFRLSEDKKGYKLYATPLSPRLMDQ 643
DB 375 GL-----SRKDLQFTKELISPYRDTNSDRGLGYVNA-----FRREISSETFDE 418
OY 644 LRATLGYQOEVEGHSITNGFDLSTRLHEISRSIIQNGCMNRYSLSVRRLDKLKTQ---A 700
DB 419 IKLANGKQVRE--GKVGIGISIQ--RPID-----GMAASIGFNYSRISINDROGNI 465
OY 701 PPEWQDLPVDFVNGKPSQEAALLAGVAVHKTVDNLVNPMBGRQRYSLVEGSSGLVSDA 760
DB 701 PPEWQDLPVDFVNGKPSQEAALLAGVAVHKTVDNLVNPMBGRQRYSLVEGSSGLVSDA 760

OC Bacteria: Proteobacteria; alpha subdivision; Caulobacter group;
OC Bacteria: Caulobacter.
OC NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Ulfenback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005865; AAK23890.1; -.
DR TIGR: CC1915; -.
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surface_Ag_1.
DR Complete proteome.
SQ SEQUENCE 769 AA; 85938 MW; 7B12A1FEB6226F2B CRC64;

Query Match 4.0%; Score 190.5; DB 16; Length 769;
Best Local Similarity 20.0%; Pred. No. 0.009;
Matches 151; Conservative 94; Mismatches 250; Indels 259; Gaps 34;

OY 250 IDLSIIRNSIGENVYIHDGEPYI--DYRAVEYREGGADDAKFTYVADVPPLIGDV 306
DB 160 VDLVEINEGAKSGVGLNFGNAEYSDNLDYIVYTESRYKILTS-----NDN 210
OY 307 FHHGKYEYTKKNLLENASAEHGFYDGRMLDRSDVILPD--NTADVSLYDTGTQYREDEV 365
DB 211 YDPRILEYDRQGLRKHYNNRGYFDFRYS--SVALAPDKNGAVYTLLEBGFKAFGRF-- 268
OY 366 FETIDPKTNQLTTPDKLPVKKRELLQQLLTVMNGEAYNLQAVRALSNDLIATRYFNMYNT 425
DB 269 --TYETELKKLDGN-----LLAQILLPRTGLYE----- 295
OY 426 EIVPEREQIONQVSFEQSSSS--RTEPAQVDESTLEPVLETVELTGDGLMDISPIE 481
DB 296 -----DERIEQATDALTFPAGAGAFVDVPRRYVPPNRETKTVDVVFQVREG----- 342
OY 482 FSASNLIDKLNVAAKARHLVMDPDRVLAINHDDGVNRSILGRISDASAVARAILPD 541
DB 343 ---PRVYVDRIIDY--GNTRL--DYVLKRELVEAGDAVNR-----VLYVD 381
OY 542 ESEN-----EVID-----LPERTALANRKTADVQSKRVPLVYFVASDKPRDQI 586
```

```
Db 382 RSKNNRRRLGFEKEVEDAPGSAPDRTSLR-----VKVEEQPTGE 422
QY 587 IGLWGS D T G T R L V T F E H N L I N R D G Y O A G A E I R L S E D K K G V K L Y A T K P L S H P L N D Q L R A 646
Db 423 L S F S A G Y S S I D K L V - L D V G I T E R N F R G R G Q N L R -----ARA 457
QY 647 T L G - Y O Q E Y F G H S T N G F D L S T R T L E H E I R S T I O N G M N R T Y S L R Y R L D K L T Q A P P E T 704
Db 458 S V G S L R Q O I D F G S E P R F -----I G R N L Y - A G V N - L Y T F R Y D L S E F A Y D T K S V 504
QY 705 M O D L P V D F V N G R K P S Q E A L L A G V A V H K - T V A D N L -----736
Db 505 G G D V R G F P L T N D S S M S L R T V R O D E V S A D S L C A S G S V S Q I L C L O R G A Y I T S L I G Y G L R 564
QY 737 -----V N P M R G Y R Q R Y S L E V G S S G L V S D A N M A I A R A G I S G V Y S F G D N A Y G S N R A -----785
Db 565 I D K R N D P I N P T R G W -----F A D L N Q D L A G V -----G D V K Y L K T E A D A G W 604
QY 786 -----H O M T G G I O A G Y I M S D N F N H P Y R L R F F A G D Q S I R G Y A H D L S P - I S D K G 834
Db 605 Y M G F T R D L V F S A T G S F - G Y I E G W G D N V R I N D R F Y R G G - T S F R G F E I A G I G P R D I S S F 661
QY 835 Y L T G G O V L A V G T A E Y N Y - E F M K D - - L R L A V F G D I G N A Y D K G F T N D P -----877
Db 662 N S M G A K L Y A I S T F E L V P T L P E O Y G I K A L F S D V G T A - - G L L D V D R O R S P G V F D P N I 718
QY 878 -----K I G A G V G R W A S P V G O V R V D A T G V K E E 905
Db 719 K D N L G L R A S A G I S I D M K S P W G P I R F D I S R I L S K E 752
```

Search completed: April 28, 2003, 16:23:06

Job time : 68 secs

